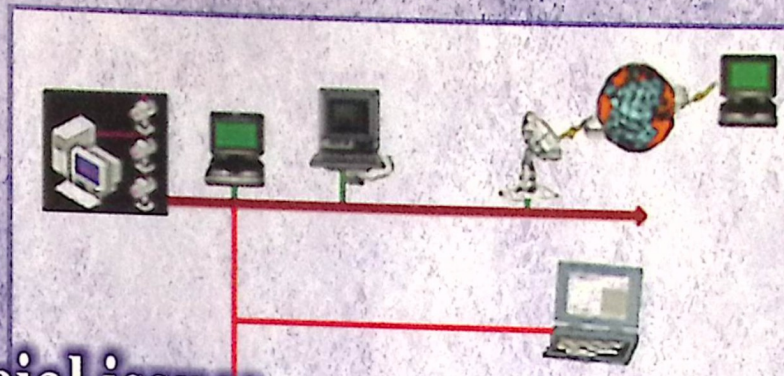
RESEARCH IN COMPUTING SCIENCE

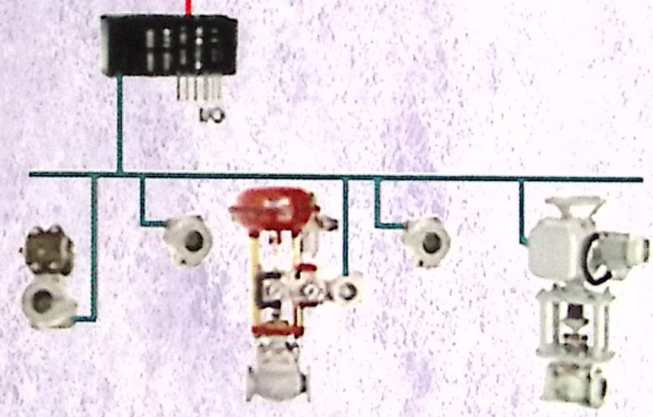
ISSN: 1870-4069



Special issue: Control, Virtual Instrumentation and Digital Systems



Luis Pastor Sánchez Fernández
Osvaldo Espinosa Sosa
(Eds.)



Vol. 24

RCS
Research in Computing Science

Special Issue: Control, Virtual Instrumentation and Digital Systems

Research in Computing Science

Series Editorial Board
Comité Editorial de la Serie

Editors-in-Chief:
Editores en Jefe

Juan Humberto Sossa Azuela (Mexico)
Gerhard Ritter (USA)
Jean Serra (France)
Ulises Cortés (Spain)

Associate Editors:
Editores Asociados

Jesús Angulo (Frane)
Jihad El-Sana (Israel)
Jesús Figueroa (Mexico)
Alexander Gelbukh (Russia)
Ioannis Kakadiaris (USA)
Serguei Levachkine (Russia)
Petros Maragos (Greece)
Julian Padget (UK)
Mateo Valero (Spain)

Editorial Coordination:
Coordinación Editorial

Blanca Miranda Valencia

Formating:
Formación

Luis Pastor Sánchez Fernández
Osvaldo Espinosa Sosa

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 24**, Noviembre, 2006. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. 04-2004-062613250000-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor Responsable: *Juan Humberto Sossa Azuela, SOAJ 560723*

***Research in Computing Science** is published by the Center for Computing Research of IPN. **Volume 24, November**, 2006. Printing 500. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research. Printed in Mexico City, Mayo, 2006, in the IPN Graphic Workshop – Publication Office.*

Volume 24
Volumen 24

Special Issue: Control, Virtual Instrumentation and Digital Systems

Volume Editors:
Editores del Volumen

Luis Pastor Sánchez Fernández
Osvaldo Espinosa Sosa

Instituto Politécnico Nacional
Centro de Investigación en Computación
México 2006

ISSN: 1870-4069

Copyright © Instituto Politécnico Nacional 2006
Copyright © by Instituto Politécnico Nacional 2006

Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal
Unidad Profesional "Adolfo López Mateos", Zacatenco
07738, México D.F., México

http.//www.ipn.mx
http://www.cic.ipn.mx

The editors and the Publisher of this journal have made their best effort in preparing this special issue, but make no warranty of any kind, expressed or implied, with regard to the information contained in this volume.

All rights reserved. No part of this publication may be reproduced, stored on a retrieval system or transmitted, in any form or by any means, including electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the Instituto Politécnico Nacional, except for personal or classroom use provided that copies bear the full citation notice provided on the first page of each paper.

Indexed in LATINDEX
Indexada en LATINDEX

Printing: 500
Tiraje: 500

Printed in México
Impreso en México

Preface

Computer sciences have experimented enormous advances in recent years, that is the reason to include in this volume of *Research in Computing Science* several contributions obtained by specialists in areas such as: Digital Signal Processing, Image Processing, Virtual Instrumentation, Digital Systems Design and Control, specialized areas that are having a high impact on contemporary societies.

We received several papers from researches of many parts of the world. We accepted only the best papers. Every paper was evaluated by three reviewers with the common objective of select the papers with the highest quality.

We want to express sincerely acknowledge to all the people that contribute with this magazine. We are specially indebted with authors, reviewers and scientific committee members for their outstanding work and advice. Finally We would also like to thank to the National Polytechnic Institute for facilities given in order to achieve the success of this special issue, which we believe, is so much important for the academic and scientific life in this country.

November, 2006

Luis Pastor Sánchez Fernández
Osvaldo Espinosa Sosa

Scientific Committee

Abelardo del Pozo	Institute of Cybernetics, Mathematics & Physics (ICIMAF). Cuba
Alberto Aguado	Institute of Cybernetics, Mathematics & Physics (ICIMAF). Cuba
Alexander Veidenbaum	University of California, Irving, USA
Alfonso Gutiérrez Aldana	Center for Computing Research, National Polytechnic Institute. México
Amaury Caballero	Florida International University. USA
Carlos Aguilar Ibañez	Center for Computing Research, National Polytechnic Institute. México
Carlos Canudas de Wit	Institut National Polytechnique de Grenoble (INPG). France
Cornelio Yáñez Márquez	Center for Computing Research, National Polytechnic Institute. México
Eduardo Quiles Cucarella	Polytechnic University of Valencia. Spain.
Elsa Rubio Espino	Center for Computing Reseach, National Polytechnic Institute. México
Emilio García Moreno	Polytechnic University of Valencia. Spain.
Humberto Ferasoli Filho	São Paulo State University - UNESP. Brazil
Jeff Meisel	Academic Accounts. National Instruments México
João E. M. Perea Martins	São Paulo State University - UNESP. Brazil
José Mireles Jr.	University of Texas at Arlington USA. Universidad Autónoma de Ciudad Juárez, México.
José de Jesús Medel Juárez	Center for Computing Research, National Polytechnic Institute. México
José G. Rodríguez	CINVESTAV. México D.F.
José Luis Verdegay	Granada University. Spain
Jose Martins Ferreira	University of Porto, Portugal
Kang Yen	Florida International University. USA
Lázaro Gorostiaga Canepa	Center of Automatization, Robotics and Technologies of the Information and the Manufacture. Spain
Luis Enrique Ramos Velasco	Autonomous University of Hidalgo State. México
Luis Miguel Fernández	Institute of Cybernetics, Mathematics & Physics (ICIMAF). Cuba
Luis Villa Vargas	Mexican Institute of Petroleum (IMP). México
Marcelo Nicoletti Franchin	São Paulo State University - UNESP. Brazil
Marco A. Ramírez Salinas	Center for Computing Research, National Polytechnic Institute. Mexico. Computer Architecture Department DAC-UPC, Barcelona, Spain
Marcos de Alba	Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM). México
Mateo Valero	Polytechnic University of Catalonia, Spain
Miguel García Silvente.	Granada University. Spain

Norberto Flores Guzmán	Mexican Institute of Petroleum (IMP). México
Oleksiy Pogrebnyak	Center for Computing Research, National Polytechnic Institute. México
Orestes Llanes Santiago	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Oscar Camacho Nieto	Center for Computing Research, National Polytechnic Institute. México
Pedro Castillo García	Centre de Recherches de Royallieu. France
Pedro Guevara López	Research Center for Applied Science and Advanced Technology (CICATA). National Polytechnic Institute. México
Ramon Beivide	University of Cantabria. Spain.
Raynitchka Tzoneva	Cape Peninsula University of Technology. Republic of South Africa
Romeo Ortega.	Ecole Supérieur d'Electricité. France
Sergio Suárez Guerra	Center for Computing Research, National Polytechnic Institute. México
Virgilio López Morales	Autonomous University of Hidalgo State. México

Table of Contents

Page

FPGA Implementation of Morphologic Primitives 1
Agustín Cruz Contreras, Juan Luis Díaz de León Santiago, Oleksiy Pogrebnyak and Pablo Manrique Ramírez

Optimized Implementation of a Pattern Classifier using Feature Set Reduction 11
Mario Aldape Pérez, Cornelio Yañez Márquez and Luis Octavio Lopez Leyva

Haar Wavelet Transform for Low Power Video Frames 21
Héctor Silva-López and Sergio Suárez Guerra

Component for Debugging of Interrupt Based Systems 31
Alain Tamayo-Fong, Luis Eduardo Leyva-del-Foyo and Pedro Mejia Álvarez

Real-time Constraints for Adaptive Digital Filters on SISO Stochastic Systems 43
Pedro Guevara López, José J. Medel Juárez and Blanca Elizabeth López Silva.

Diagnostic of Three-Phase Induction Machine Based on Neural Networks 55
Rodrigo López Cárdenas, Luis Pastor Sánchez Fernández and Beatriz Adriana Jaime Fonseca

Proposal for a Nanosatellite Microwave Communication System, to Achieve Energy Savings, High Capacity Data Transmission, and Position Control Feedback 65
Ante Salcedo and Esaú Vicente-Vivas

A Method to Classify Digital Clocks Employing Image Segmentation and Associative Memories 75
José Luis Oropeza and Sergio Suárez Guerra

Low Cost Multi-Channel Data Acquisition System with Graphical HMI 85
Antonio Hernández, Oscar Camacho Nieto, Cornelio Yañez Márquez and Osvaldo Espinosa

Patterns Recognition for Shrimp Aquaculture 95
Luis P. Sánchez, José Juan Carbajal Hernández and Oleksiy Pogrebnyak

Real-Time Constraints for Adaptive Digital Filters on MIMO Stochastic Systems 105
José J. Medel Juárez, Pedro Guevara López and Asdrúbal López Chau

Steps Towards the Development of a Portable and Cost-Effective System for Human Resources Training in Small Satellite Technology 117
Esaú Vicente-Vivas, Israel Nava, Victor Lomas, Jorge Morales, E. A. Jiménez, Z. Carrizales and F. Gómez

Teaching the Design of Takagi-Sugeno Fuzzy Controllers for an Under-Graduate Course 131
Rajesh R., Sureshkumar T., Kaimal M.R. and Suresh K.K

From Virtual Instruments to Virtual Laboratory 141
Luis Miguel Fernández , Guillermo Alvarez, Miguel Machirán, Ileana Sánchez and Luis Alejandro Sánchez

Environmental Noise Monitoring. 1st. Stage: Aircrafts Noise Patterns Recognition 151
Luis P. Sánchez, Lucrecia Pérez Echazabal and Luis Alejandro Sánchez Pérez

A Fundamental Analysis based Neuro-fuzzy Approach for Modeling the Effects of Crude Oil Ending Stocks Variations on Oil Price..... 161
Ali Ghaffari, Mansour Nikkhah Bahrami and Seyed Mahdiar Ghaffari

Modeling and Control of a Sugar Cane Process 171
Lázaro Gorostiaga Canepa, Yovany González Martínez and José Ramón Perán

Implementing RTSJ-optimized Java Processor for Real-Time Applications 179
Zhilei Chai, Wenke Zhao and Wenbo Xu

Estimation of Symmetric Positive Definite of Matrices using Evolution Strategy for Differential Games 189
Héctor Vargas, Vittorio Zanella, Vladimir Alexandrov, Monica López and Lorna Rosas

Self-Tuning Neural Controller and its Application to a Non Linear System 199
Alfonso Gómez, Alfonso Noriega, Alberto Aguado and Roberto Salas

Speech Recognition using Based-Phone, Syllables, Triphones and Words on a Digit Corpus in Spanish Language 213
José Luis Oropeza-Rodríguez and Sergio Suárez Guerra

Real-Time Fuzzy Digital Filter: Basic Concepts 227
Juan Carlos García Infante, José J. Medel Juárez and Pedro Guevara López

Author Index 239

FPGA Implementation of Morphologic Primitives

Agustín Cruz Contreras[1], Juan Luís Díaz de León Santiago[2]
Oleksiy Pogrebnyak[3] and Pablo Manrique Ramírez[4]

Instituto Politecnico Nacional, Centro de Investigacion en Computacion, Ave. Juan de Dios Batiz S/N, C.P. 07738, Mexico, D.F., Mexico
[1]aguscruz@prodigy.net.mx, [2]jdiaz@cic.ipn.mx, [3]olek@pollux.cic.ipn.mx, [4]pmanriq@cic.ipn.mx

Abstract. Specific hardware architecture for morphologic video image processing is proposed. The developed system processes high resolution black and white images in real time. The principle of the presented architecture is based on the local image data processing at the current moment. The image rows needed to cover the required pixel vicinity are formed by the use of shift registers. At the processing stage the operations of erosion and dilatation are performed using logic operators AND and OR. The hardware implemented in reconfigurable FPGA with a density of 100K logic gates and maximal frequency of 200 MHz.

Keywords: FPGA, morphologic image processing, edge detection

1 Introduction

Nowadays near real-time image processing systems are widely used in different applications, such as industrial monitoring, traffic control, vehicle inspection, robotic vision, security and vigilance, medical image processing. The data in such applications are often characterized by the presence of dynamic environment that requires the vision system must produce a fast response. Attempts to fit the requirements imposed for real-time image processing systems ususally are based on parallel type architectures: digital signal processors (DSP), PC-clusters, Single Instruction Multiple Data systems (SIMD), for example, 3DNow! from AMD y SSE (Streaming SIMD Extensions) from Intel, IMAP-VISION PCI board [1] that uses SIMD architecture and 16 processors.

The parallelism offer a lot of benefits, but the implementation of the processing algorithms is can not be done in one step. Under the requirement of a immediate response from the vision systems various alternatives were developed. Besides, the specific architectures plan algorithm implementation as a 100% hardware solution that provides dedicated resources for the algorithm execution. This alternative is usually implemented as Application Specific Integrated Circuits (ASIC), and FPGA can be used in case when the algorithm complexity permits it.

From the other hand, mathematical morphology theory arisen in 60's. Morphologic image processing techniques permit to extract easily image components useful for image region representation and description, such as edges, skeletons and convex forms. The mathematical morphology language is the theory of groups that represent

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 1-10

the forms and shapes of the image objects. The fundamental morphologic operation is nonlinear and consists of the image compare to a simple geometric pattern named structure element. The mayor advantages of the morphologic image processing are: low computational complexity caused by logic operators used; the processed image preserves its geometric features; two basic primitives, namely dilation and erosion, in combination to themselves form all operators.

In this paper, we present a specific architecture based on the data flow that implements a morphologic image processing offering the following advantages: it is not necessary to acquire the all image to start processing; the introduced time delay is determinated by the buffer length (2 rows and three pixels for each operation); once the buffer is filled out the delay is zero; for its simplicity, the architecture can be implemented in FPGA.

2 Edge Detecting Using Basic Morphologic Operators

The operation fundamental of basic morphologic operations, dilation and erosion , one can found in the literature, for example, in [2]. The binary dilation can be calculated as a sum of Minkowsky $A \oplus B = \{x = a + b \in X | a \in A \wedge b \in B\}$, as a union of translation and from a geometric algorithm that can be formulated as follows. Let $A \subset X$ *and* $B \subset X$ are two sets, thus from Theorem 3.8 in [2] it follows

$$A \oplus B = \left\{ x \middle| \left(B^-\right)_x \cap A \neq \phi \right\}, \tag{1}$$

where B^- denotes reflected set B, $B^- = \{-b | b \in B\}$, $\left(B^-\right)_x$ denotes translated set B^-, $(B)_x = \{y = b + x | b \in B\}$, ϕ is an empty set. The geometric algorithm to dilate the set $A \subseteq X$ by a structural element $B \subseteq X$ can be described step bye step as follows [2]:

1. Choose a point $x \in X$
2. Calculate B^-, the reflection of the structural element B
3. Find $\left(B^-\right)_x$, the translation of B^- by x
4. Realize the intersection $\left(B^-\right)_x \cap A$
5. If the set $\left(B^-\right)_x \cap A \neq \phi$, thus $x \in A \oplus B$
6. Steps 1 to 5 are repeated for all the space X points.

The binary erosion can be found as a rest of Minkowsky [2]: $A \ominus B = \{x \in X | x + b \in A, \forall b \in B\}$. Besides, the erosion can be defined as a geometric algorithm, in the same way as it was defined for the dilatation [2]. Theorem 4.2 [2] states that

$$A \ominus B = \left\{ x \middle| (B)_x \subseteq A \right\}. \tag{2}$$

This result provide a geometric method to find the erosion of a set $A \subseteq X$ by a structural element $B \subseteq X$. The algorithm consists of the following steps [2]:

1. Choose a point $x \in X$.
2. Find $(B)_x$, the translation of B by x.
3. If $(B)_x$ is a subset of A, thus $x \in A \Theta B$.
4. Repeat steps 1-3 for all points of the space X. Finally, the set $A \Theta B$ will be found.

The opening operator [2] of the set A by the structural element B is denoted as $A \circ B$ and is defined as $A \circ B = (A \Theta B) \oplus B$. The effects of the opening on an image can be resumed in four aspects: it eliminates islands of the size smaller tan of the structural element; it eliminates picks o valley thinner than the structural element; it breaks isthmuses whose width is smaller than the structural element diameter; it smooth the image convex edges.

The closing operator [2] of the set A by the structural element B is denoted as $A \bullet B$ and is defined as $A \bullet B = (A \oplus B) \Theta B$. The effects of the closing are: it fills lakes or holes of the size lower than of the structural element; it fills cracks or gulfs thinner than the structural element; it fuses straits whose width is smaller of the diameter of the structural element; it planes the concave image edges, stuffing breaks.

Similarly to the operators of the erosion and dilatation, it is possible to design the closing operator from the opening operator and vice versa using the property of the duality between the erosion and dilation operators. Additionally, these operators share the property of *idempotence*, which means that once the opening or closing is performed, the repetition of such a process does not have any effect on the working set. In other words, the repetition of the opening or closing with the same structural element does not alter the result.

The Theorem 3.5 form [2] states that given the A, B subsets of X, the inner border is equal to

$$\Gamma(A) = A - (A \Theta B). \tag{3}$$

Fig 1 illustrates the mechanism of edge extraction: (a) shows the original image, (b) shows the structural element, (c) presents the erosion of A by B, (d) shows the found edge $\Gamma(A) = A - (A \Theta B)$.

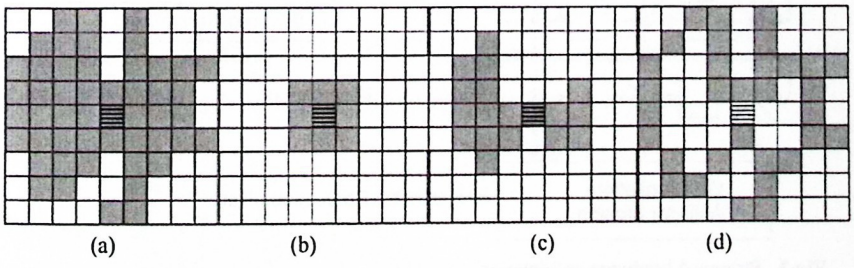


Fig 1. Edge extraction: (a) original image, (b) structural element, (c) erosion of A by B, (d) edges of A

Considering all values as one for non-zero pixels and zero for the background, the edge extraction operation can be implemented by means of logic operators, this is, for $-(A \Theta B)$ it can be used AND and an inverter to complement the operation, for $\Gamma(A) = A - (A \Theta B)$ it can be used XOR between A and the inverter output.

3 Proposed Hardware Architecture and Implementation

The before mentioned basic morphologic operations (erosion and dilatation) can be realized locally in the vicinity of the current pixel. For the vicinity of one pixel nine pixels (3 x 3) are required in three rows and three columns to define the desired structural element. In this manner the processing can start without the necessity to have the entire image. In the video signal the data are transmitted in a serial form, say the image arrives pixel by pixel. The image is formed from left to right and from top to bottom. This situation permits us to see that from three received image rows it is possible to start image processing. This way the image can be processed as a signal.

Fig. 2 shows the proponed architecture for the operations of erosion and dilation, with a structural element size of 3 x 3 pixels. This architecture consists of two shift registers having the longitude in pixels equal to one video row, and the third raw may have the longitude of three first pixels that it is sufficient to start processing.

With the first three pixels of three rows the processing window is formed to obtain the first processed pixel. To perform the dilatation OR is applied to the current pixel and its vicinity; in case of the erosion the operator AND is applied. The output of the gates thus corresponds to the processed signal. To process next pixel, a shift in the registers is performed, then the morphologic operators are applied. This way the processing is repeated until entire video frame image is passed through the registers.

As it can be evaluated, the processing time is equal to the time of the logic gate trough pass plus the delay corresponding to the duration of dos video rows and three pixels of the third one. Once the registers are filled out the processing time is zero.

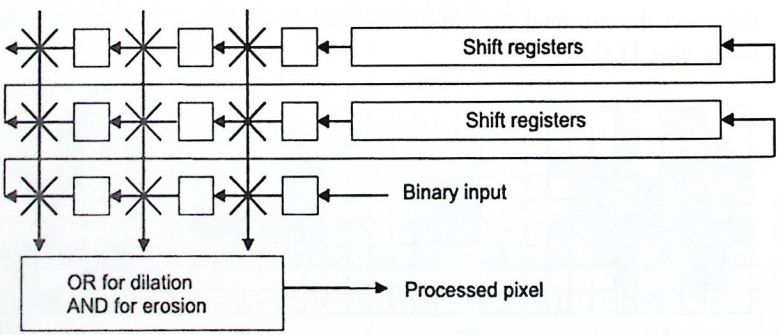


Fig 2. Proposed hardware architecture

At the first sight, the presented idea is simple and functional, but it is not very obvious the way to implement the proposed architecture in commercial electronic devices, given that the price of the design using a specific circuit may be very high for this purpose. The required device for the proposed architecture is the shift register that is common used and widely represented at market. The required shift register quantity depends on the length of the image row. For high quality video the image size often is 640x480 thus to implement the proposed architecture one requires 2 x 640 +3=1283 flip-flops.

FPGA are flexible electronic devices with high density of gates and fast propagation. Revising the possibilities of their inner architecture one can see that using only the flip-flops imply to waste the mayor percentage of the device resource. This situation causes that the implementation results expensive. Besides, we found the XILINX FPGAs predispose in the architecture of their devices the possibility to implement large shift registers. The XILINX devices such as SPARTAN II, SPARTAN III and VIRTEX offer the possibility to implement the shift registers without supersaturate the inner flip-flops of the device. For this purpose the look-up tables (LUT) can be used [3], [4]. In this manner all resources of the FPGA can be employed.

XILINX's design software include various libraries for circuits, among them are the primitives for the shift registers of 16 bits implemented by LUT. Using the XILINX primitive for 16 bit shift register, the frame for the shift register of 640 bits can be designed by means of 40 cascades of 16 bit shift registers. With this frame one row of the image pixels can be stored. Once the 640 bit register line is designed, it can be invoked the number of times it is necessary. Having the line of the shift register of 640 bits and discrete flip-flops, the proposed architecture for the implementation of the erosion and dilation primitives is conformed.

The primitives for the erosion and dilation have as a formative element the shift register of 16 bits implemented by LUT. To form the frame, various registers of this type are connected in cascade (see Fig. 3).

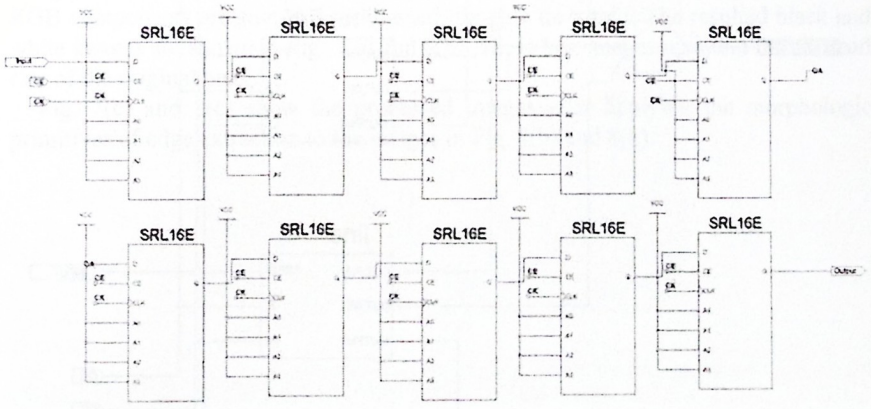


Fig. 3. Block of ten shift registers of 16 bits

Fig. 4 shows the block of 4 bits that is used at the end of the line to implement the operations OR and AND.

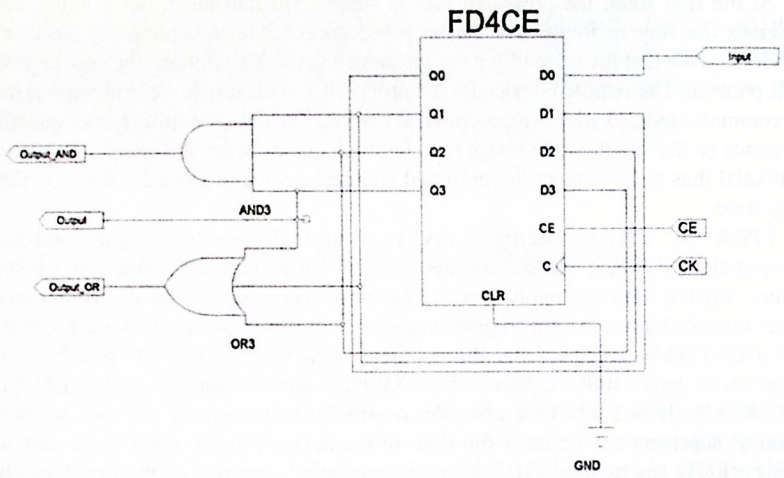


Fig. 4. Block at frame end

The basic primitive of erosion and dilation is shown in Fig. 5. This primitive faithfully implements the proposed architecture. It is formed by two frames of 640 bits and the logic gates OR (erosion) and AND (dilatation).

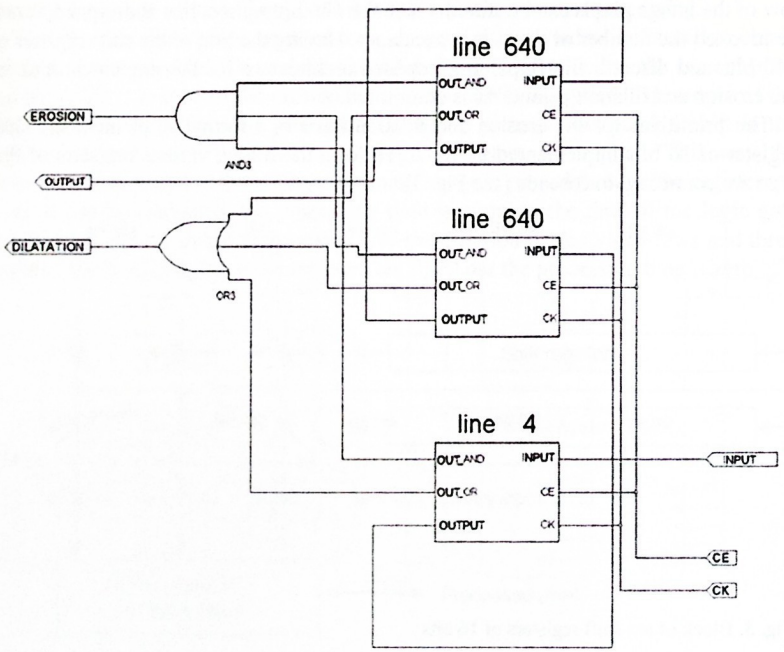


Fig. 5. Primitive for erosion and dilation

The primitive for edge extraction is shown in Fig. 6. It integrates the erosion-dilation primitive, one delay, one inverter and one XOR to implement the equation (3).

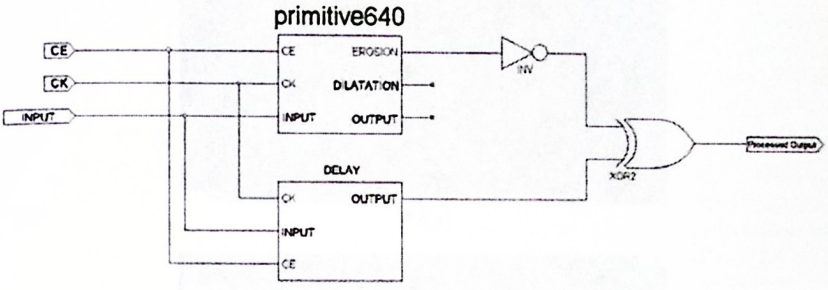
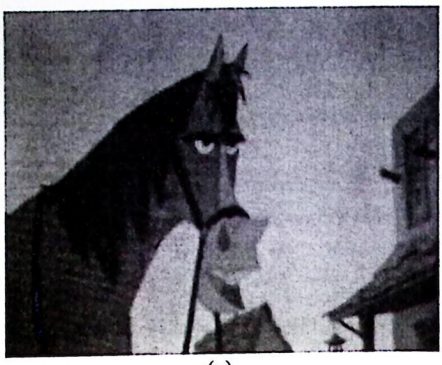


Fig. 6. Edge primitive.

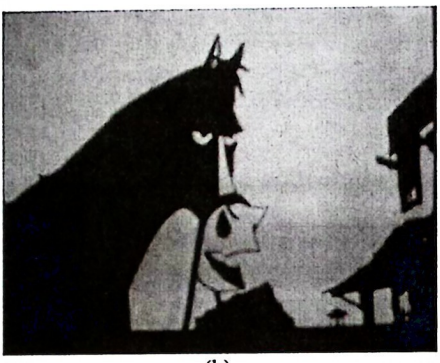
4 Example of edge detection application

In this section, we present the results related to image edge extraction as an example of the morphologic processing. Fig. 7 and 8 show the original and processed images that are the photos captured from PC monitors. The original color images (see Fig. 7(a) and Fig. 8(a)) of 640x480 pixels are the frames of video stream of 60fps. First, the color images are converted to their black/white representation. For this purpose the video data are got form DVI connector of a VGA monitor and are converted to the RGB format using a hardware decoder, for example, SIL161B, SIL163B from Silicon Image or TFP101, TFP401 from Texas Instruments can be used [5]-[7]. Next, the RGB components are summed and passed trough a threshold. The resulted black and white images are shown in Fig. 7(b) and 8(b). These b/w images maintain the size and rate of the original ones.

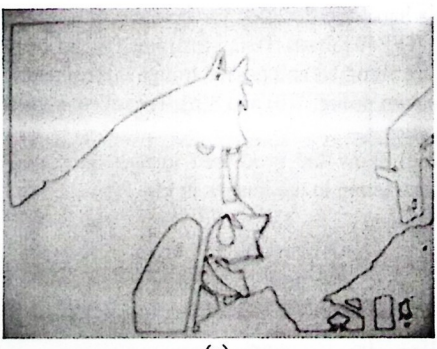
Fig. 7(c) and 8(c) show the processed images alter applying the morphologic primitives of edge extraction to the images in Fig. 7(b) and 8(b).



(a)



(b)

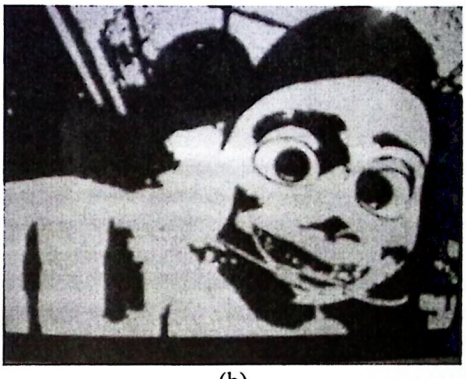


(c)

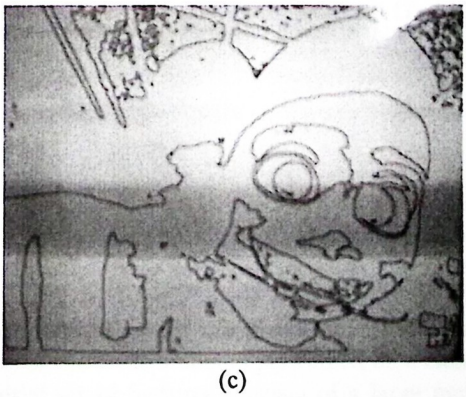
Fig. 7. (a) color image, (b) black and white image, (c) edge image.



(a)



(b)



(c)

Fig. 8. (a) color image, (b) black and white image, (c) edge image.

5 Conclusions

We have presented the hardware architecture that permits to process the morphologic primitives with a minimal time delay. The introduced delay directly related to the structural element size and for the size of 3 x 3 pixels the delay corresponds to the duration of three rows of image. The employed FPGA has the density of 100,000 gates and system frequency of 200Mhz. The use of FPGA look-up tables as shift registers benefits in the full usage of FPGA resources and permits to implement in average 40 morphologic primitives. The designed FPGA based system processes satisfactory the images with the resolution of 640 x 480 pixels and rate of 60 frames per second. Using the developed frames for the morphologic primitives and software scheme editor the designed system is flexible and can be used in various application of morphologic image processing.

References

1. Moler, van der, M.W., Kyo, S., "Documentation for the IMAP-VISION image processing card and the 1DC language", NEC incubation Center, 1997
2. Díaz de León Santiago J., Yáñez Márquez C. (2003), *Introducción a la morfología matemática de conjuntos*, IPN, UNAM, FCE, México, D.F.
3. Spartan-II 2.5V FPGA Family: Complete Data Sheet, Xilinx
4. Using Look-Up Tables as Shift Registers (SRL16) in Spartan-3 Generation FPGAs, Xilinx
5. TFP101, TFP101A Design Notes, Texas Instruments.
6. Sil161B PanelLink Receiver Datasheet, 2001, Silicon Image, Inc.
7. TFP401, TFP401A, TI PanelBus DIGITAL RECEIVER, Texas Instruments.

Optimized Implementation of a Pattern Classifier using Feature Set Reduction

M. Aldape-Pérez, C. Yáñez-Márquez, and L.O. López Leyva

Center for Computing Research, CIC
National Polytechnic Institute, IPN
Mexico City, Mexico
maldapeb05@sagitario.cic.ipn.mx; cyanez@cic.ipn.mx
WWW home page: http://www.cornelio.org.mx

Abstract. Performance in most pattern classifiers is improved when redundant or irrelevant features are removed, however, this is mainly achieved by high demanding computational methods or successive classifiers construction. In this paper new results on coherent feature selection, obtained by applying the Hybrid Classification and Masking technique (HCM), are presented. Previous results showed that the HCM algorithm proves to be a feasible way to get an optimal subset of features represented by a mask value. This method identifies irrelevant or redundant features for classification purposes; an optimal subset of features allows register size optimization which clearly contributes to significant power savings. Moreover classifier accuracy of this new approach has been preserved or even improved in comparison to other experimental results formerly shown. Promising experimental results suggest that the HCM algorithm is an appropriate alternative for optimized pattern classifier hardware implementation.

Keywords: Feature Selection, Pattern Classifier, Masking Techniques, Supervised Learning

1 Introduction

Pattern recognition has existed for many years in a wide range of human activity, however, the general pattern recognition problem can be stated in the following form: Given a collection of objects belonging to a predefined set of classes and a set of measurements on these objects, identify the class of membership of each of these objects by a suitable analysis of the measurements (features) [1]. Although features are functions of the measurements performed on a class of objects, most of the times, the initial set of features consists of a large number of potential attributes that constitute an obstacle not only to the accuracy but to the efficiency of algorithms. In countless situations, it is a complicated task to find proper features for all patterns in a class (especially when intra-class variance is very high) [2]. In order to overcome this limitation, multiclassifier approach arises. Some of the most frequently used are: multiple models [3], [4], [5]; combining classifiers [6]; and classifier ensembles [7] among others. On behalf of these

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 11-20

novel approaches a remarkable thing to mention is the notable classification rate achieved; nonetheless these methodologies lack of criterions that help to ignore redundant or irrelevant information.

In this paper an original method for pattern classifier hardware implementation, whose basic operation rests on a Hybrid Associative Memory and the HCM algorithm for optimality criterion evaluation, is presented.

Each obtained mask value, represents a different subset of features. It is to be said that the best mask value is the one that indicates the smallest subset of features, and hence to permit register size optimization for hardware implementation purposes.

From a register transfer level perspective, the smaller the register size is, the better architecture alternative is obtained.

In the following section, a brief description of *HCM* foundations is presented. In Section 3, experimental results are shown over several different data sets. The masking approach advantages will be discussed in section 4, and a short conclusion follows in Section 5.

2 Associative Memories

An associative memory M is a system that relates input patterns and output patterns as follows: $x \longrightarrow \boxed{\mathrm{M}} \longrightarrow y$ with x and y, respectively, the input and output pattern vectors. Each input vector forms an association with its corresponding output vector. For each k integer and positive, the corresponding association will be denoted as: (x^k, y^k). Associative memory M is represented by a matrix whose ij-th component is m_{ij} [9]. Memory M is generated from an *a priori* finite set of known associations, called the fundamental set of associations. If μ is an index, the fundamental set is represented as: $\{(x^\mu, y^\mu) \mid \mu = 1, 2, ..., p\}$ with p as the cardinality of the set. The patterns that form the fundamental set are called fundamental patterns. If it holds that $x^\mu = y^\mu \; \forall \mu \in \{1, 2, ..., p\}$ M is auto-associative, otherwise it is heteroassociative; in this case it is possible to establish that $\exists \mu \in \{1, 2, ..., p\}$ for which $x^\mu \neq y^\mu$. If we consider the fundamental set of patterns $\{(x^\mu, y^\mu) \mid \mu = 1, 2, ..., p\}$where n and m are the dimensions of the input patterns and output patterns, respectively, it is said that $x^\mu \in A^n$, $A = \{0, 1\}$ and $y^\mu \in A^m$. Then the j-th component of an input pattern is $x_j^\mu \in A$. Analogously, the j-th component of an output pattern is represented as $y_j^\mu \in A$. Therefore the fundamental input and output patterns are represented as follows:

$$x^\mu = \begin{pmatrix} x_1^\mu \\ x_2^\mu \\ \vdots \\ x_n^\mu \end{pmatrix} \in A^n \qquad y^\mu = \begin{pmatrix} y_1^\mu \\ y_2^\mu \\ \vdots \\ y_m^\mu \end{pmatrix} \in A^m$$

2.1 The Steinbuch's Lernmatrix

Lernmatrix is a heteroassociative memory that can easily work as a binary pattern classifier if output patterns are appropriately chosen [13]. Typically it ac-

cepts binary patterns suchlike $\mathbf{x}^\mu \in A^n$, $A = \{0,1\}$ as input and returns binary patterns suchlike $\mathbf{y}^\mu \in A^m$ as output; it is worth pointing out that there are m different classes, each one coded by a simple rule: class $k \in \{1,2,...,m\}$ will be represented by a column vector which components will be assigned by $y_k^\mu = 1$, so $y_j^\mu = 0$ for $j = 1,2...,k-1,k+1,...m$.

The following matrix will keep the pattern association values after the Learning Phase for the Steinbuch's *Lernmatrix* is done:

	x_1^μ	x_2^μ	$\cdots$	x_j^μ	$\cdots$	x_n^μ
y_1^μ	m_{11}	m_{12}	$\cdots$	m_{1j}	$\cdots$	m_{1n}
y_2^μ	m_{21}	m_{22}	$\cdots$	m_{2j}	$\cdots$	m_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
y_i^μ	m_{i1}	m_{i2}	$\cdots$	m_{ij}	$\cdots$	m_{in}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
y_m^μ	m_{m1}	m_{m2}	$\cdots$	m_{mj}	$\cdots$	m_{mn}

(1)

Each one of the m_{ij} components of M is initialized with zero and will be modified by the following rule: $m_{ij} = m_{ij} + \Delta m_{ij}$ where:

$$\Delta m_{ij} = \begin{cases} +\epsilon \text{ if } y_i^\mu = 1 = x_j^\mu \\ -\epsilon \text{ if } y_i^\mu = 1 \text{ and } x_j^\mu = 0 \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

and ϵ a positive constant, previously chosen.

The Recalling Phase for the Steinbuch's *Lernmatrix* consists of finding the class which an input pattern $\mathbf{x}^\omega \in A^n$ belongs to. Finding the class means getting $\mathbf{y}^\omega \in A^m$ that corresponds to $\mathbf{x}^\omega$; accordingly to the construction method of all $\mathbf{y}^\mu$, the class should be obtained without ambiguity. The i-th component of y_i^ω is obtained according to the following rule, where $\vee$ is the *maximum* operator:

$$y_i^\omega = \begin{cases} 1 \text{ if } \sum_{j=1}^n m_{ij}.x_j^\omega = \bigvee_{h=1}^m \left[\sum_{j=1}^n m_{hj}.x_j^\omega\right] \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

2.2 Linear Associator.

It is worth pointing out that James A. Anderson and Teuvo Kohonen obtained amazingly similar results known nowadays as *Linear Associator*. Lets consider the fundamental set as $\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1,2,...,p\}$ with

$$\mathbf{x}^\mu = \begin{pmatrix} x_1^\mu \\ x_2^\mu \\ \vdots \\ x_n^\mu \end{pmatrix} \in A^n \quad \text{y} \quad \mathbf{y}^\mu = \begin{pmatrix} y_1^\mu \\ y_2^\mu \\ \vdots \\ y_m^\mu \end{pmatrix} \in A^m$$

The *Learning Phase* is done in two stages.

1. Consider each one of the p associations $(\mathbf{x}^\mu, \mathbf{y}^\mu)$, so an $m \times n$ matrix is obtained by $\mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^t$

$$\mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^t = \begin{pmatrix} y_1^\mu \\ y_2^\mu \\ \vdots \\ y_m^\mu \end{pmatrix} \cdot (x_1^\mu, x_2^\mu, ..., x_n^\mu) = \begin{pmatrix} y_1^\mu x_1^\mu & y_1^\mu x_2^\mu & \cdots & y_1^\mu x_j^\mu & \cdots & y_1^\mu x_n^\mu \\ y_2^\mu x_1^\mu & y_2^\mu x_2^\mu & \cdots & y_2^\mu x_j^\mu & \cdots & y_2^\mu x_n^\mu \\ \vdots & \vdots & & \vdots & & \vdots \\ y_i^\mu x_1^\mu & y_i^\mu x_2^\mu & \cdots & y_i^\mu x_j^\mu & \cdots & y_i^\mu x_n^\mu \\ \vdots & \vdots & & \vdots & & \vdots \\ y_m^\mu x_1^\mu & y_m^\mu x_2^\mu & \cdots & y_m^\mu x_j^\mu & \cdots & y_m^\mu x_n^\mu \end{pmatrix} \tag{4}$$

2. M memory is obtained by adding all the p matrices

$$\mathbf{M} = \sum_{\mu=1}^{p} \mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^t = [m_{ij}]_{m \times n} \tag{5}$$

in this way the ij-th component of M memory is expressed as:

$$m_{ij} = \sum_{\mu=1}^{p} y_i^\mu x_j^\mu \tag{6}$$

The *Recalling Phase* for the *Linear Associator* is done by operating the M memory with an input pattern $\mathbf{x}^\omega$, where $\omega \in \{1, 2, ..., p\}$; operate $\mathbf{M} \cdot \mathbf{x}^\omega$ as follows:

$$\mathbf{M} \cdot \mathbf{x}^\omega = \left[\sum_{\mu=1}^{p} \mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^t \right] \cdot \mathbf{x}^\omega \tag{7}$$

$$\mathbf{M} \cdot \mathbf{x}^\omega = \mathbf{y}^\omega \cdot \left[(\mathbf{x}^\omega)^t \cdot \mathbf{x}^\omega \right] + \sum_{\mu \neq \omega} \mathbf{y}^\mu \cdot \left[(\mathbf{x}^\mu)^t \cdot \mathbf{x}^\omega \right] \tag{8}$$

Expression 8 lets us know about which restrictions have to be observed thus perfect recalling is achieved. These restrictions are expressed as:

$$(\mathbf{x}^\mu)^t \cdot \mathbf{x}^\omega = \begin{cases} 1 \text{ if } \mu = \omega \\ \\ 0 \text{ if } \mu \neq \omega \end{cases} \tag{9}$$

If condition 9 is met, then a perfect recalling is expected. So 8 is expressed as:

$$\mathbf{M} \cdot \mathbf{x}^\omega = \mathbf{y}^\omega.$$

2.3 Hybrid Classification and Masking approach

As it was said *Hybrid Classification and Masking* technique (HCM) are presented as a new feature selection approach to provide a mask that identifies the optimal subset of features, the best mask value is the one that indicates the smallest subset of features, and hence to permit register size optimization for hardware implementation purposes. In order to explain how optimal mask is found, some definitions are required.

Definition 1. *Let* $\mathbf{f}$ *be the number of features from the original set of data.*

Definition 2. *Let* $\mathbf{r}$ *be an index where* $r \in \{1, 2, ..., (2^f - 1)\}$

Definition 3. *Let* $\mathbf{e}^r$ *be a masking vector of size* n *represented as:*

$$\mathbf{e}^r = \begin{pmatrix} e_1^r \\ e_2^r \\ \vdots \\ e_n^r \end{pmatrix} \in B^n \tag{10}$$

where $B = \{0, 1\}$

Definition 4. *Let* $\dashv$ *be a new operation called IntToVector which takes* $r \in \{1, 2, ..., (2^f - 1)\}$ *and returns a column vector* $\mathbf{e}^r$ *with* r *value expressed in its binary form. From a register transfer level perspective (RTL) this can be expressed as* $bin(r) \rightarrow [\mathbf{e}^r]$. *For example: If* $r = 11$ *then* $\dashv \mathbf{e}^r$ *returns a column vector with* r *value in its binary form so the obtained vector is:*

$$\mathbf{e}^{11} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

where e_n^r is the Least Significant Bit (LSB)

Definition 5. *Let* $\|$ *be a new operation called MagVector which takes a column vector* $\mathbf{e}^r$ *of size* n *and returns an integer and positive value according to the following rule:*

$$\| \mathbf{e}^r = \sum_{j=1}^{n} \left(e_j^r \wedge 1 \right) \tag{11}$$

Where $\wedge$ is the logical AND operator.

Another relevant thing to mention is that the *Recalling Phase* is dramatically different from the previous models; it is carried out by the following rule:

$$y_i^\mu = \begin{cases} 1 \text{ if } \sum_{j=1}^{n} m_{ij} \cdot (x_j^\mu . e_j^r) = \bigvee_{h=1}^{m} \left[\sum_{j=1}^{n} m_{hj} \cdot (x_j^\mu . e_j^r) \right] \\ 0 \text{ otherwise} \end{cases} \tag{12}$$

where $\mu \in \{1, 2, ..., p\}$ and $r \in \{1, 2, ..., (2^f - 1)\}$

It is said that the *Recalling Phase* is dramatically different from the previous models because a masking vector e^r of size n masks each input vector x^μ of size n. This is where the masking technique comes into view. Using the previous definitions and the clear advantages that inherit from the *Hybrid Associative Classifier* model, it is possible to enunciate The HCM algorithm.

1. Let n be the dimension of each input pattern in the fundamental set, grouped in m different classes.
2. Each one of the input patterns belongs to a k class, $k \in \{1, 2, ..., m\}$, represented by a column vector which components will be assigned by $y_k^\mu = 1$, so $y_j^\mu = 0$ for $j = 1, 2..., k-1, k+1, ...m$.
3. Create a classifier using expression 4, 5 and 6.
4. Use the *IntToVector* operator to get the r-th masking vector as in expression 10
5. The recalling phase is carried out according to expression 12 so an r-th accuracy parameter is obtained
6. Store both parameters (the r-th accuracy parameter and the r-th masking vector) so feature selection can be evaluated in step 8
7. Compare the r-th accuracy parameter with the $(r-1)$-th accuracy parameter. The best accuracy value is stored thus accuracy improvements are achieved with each iteration
8. The same applies to the r-th masking vector. Feature selection can be evaluated using expression 11. So the smaller this number is, a better mask is obtained.
9. The new subset of features is obtained by a mask value represented by a column vector, where accuracy and feature selection are optimal

3 Experimental results

Throughout the experimental phase, two databases taken from the UCI Machine Learning Database Repository (http://www.ics.uci.edu/~mlearn) were included. The main characteristics of these data sets have been summarized in Table 1.

Contraceptive Method Choice and Australian Credit Approval databases were chosen because both data sets differ not only in features number, but in number of patterns and classes. The experiments have been carried out as follows: In the first (Learning) phase, the same number of input vectors for each class was randomly taken, which means that a balanced classifier is guaranteed. Particularly four vectors were taken from each class to get a total of twelve input vectors for the Contraceptive Method Choice and eight input vectors for

	CMC	Credit Approval
Number of Classes	3	2
Number of Patterns	1473	690
Original Set Size	9	14
Optimized Subset Size	**4**	**10**
Feature Optimization	**44%**	**71%**
Original Register Size	130 bits	360 bits
Optimized Register Size	**78 bits**	**120 bits**
Register Size Optimization	**40%**	**66%**

Table 1. Optimization Results of the data sets used

	1-NN	3-NN	C-Means	CHAT	HCM
CMC	37.24%	38.11%	31.84%	55.25%	**58.06%**
Credit Approval	58.33%	54.88%	50.81%	59.09%	**85.51%**

Table 2. Classification Rate of 5 different algorithms

the Australian Credit Approval. In the second (Recalling) phase the whole data set $\mathbf{x}^{\mu}$ was classified for each mask vector $\mathbf{e}^{r}$, in this way classifier accuracy for every mask was evaluated. The optimal mask was obtained by comparing the two main parameters (classifier accuracy and mask optimality). The same experimental procedure was carried out for each one of the two databases.

In order to estimate the HCM optimal mask efficiency, the developed method was tested by performing the following experiment several times. In the Learning Phase, the same number of input vectors for each class was randomly taken, conversely to the previous Recalling Phase, we conducted the classification process as a bi-class problem so a coherent comparison between HCM results and previous researchers' results can be done [16]. It is to be said that the results on each data set have been averaged over twenty experiments; all conducted using the same criterion. (summarized in Table 2.)

4 Results Analysis

As it is shown in Table 1 the original number of features for each database, CMC and Australian Credit Approval, is 9 and 14 respectively. After HCM algorithm performs a coherent feature selection, represented by a mask value, optimized subsets are obtained. Feature optimization for CMC database is close to 44% which means that only 56% is relevant information for classification purposes. While using near half of the initial features HCM algorithm improves classification rate over the other classification methodologies considered, consistent results are shown in Table 2.

Even more representative results of the HCM algorithm effectiveness appear with Australian Credit Approval database. Feature optimization for this database is close to 71% which means that only 29% of the initial set of features

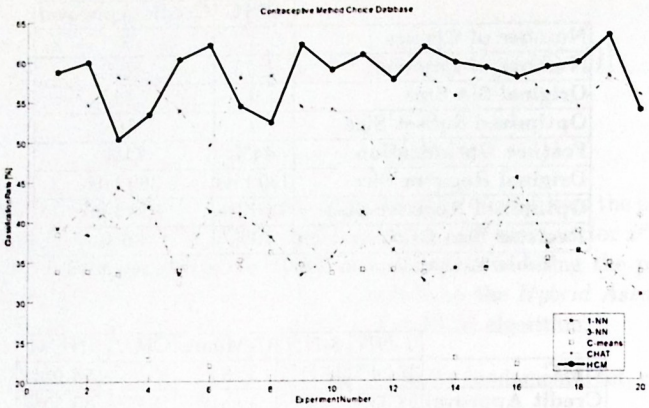


Fig. 1. Contraceptive Method Choice Database Classification Rate.

increases class separation. Whenever an optimal subset of features is obtained, astonishing classification rate will be achieved. Higher predictive accuracy can often be obtained when data dimensionality is reduced.

There are two remarkable facts to be taken into consideration. The former concerns about the number of features that was optimized for each database; 4 out of 9 features were coherently selected for the Contraceptive Method Choice database and 10 out of 14 features were coherently selected for the Australian Credit Approval database. The latter is that the HCM classification rate maintains higher predictive accuracy over the other classification methodologies considered.

5 Conclusions

In this paper an original method for optimized pattern classifier hardware implementation, whose basic operation rests on the HCM algorithm for optimality criterion evaluation, is presented. Experimental results have shown that this algorithm is an efficient way to get a mask value which represents the optimal subset of features. While maintaining the discriminatory information necessary for classifier accuracy improvement, these optimized subsets allow considerable register size reduction for hardware implementation purposes. A remarkable thing to mention is that feature optimization not only reduces data dimensionality, which represents an important reduction on hardware resources, it also means significant power savings on ASIC implementations due to register size reduction. After HCM algorithm performs a coherent feature selection, promising results have shown that accuracy has been improved in comparison to the other methodologies considered.

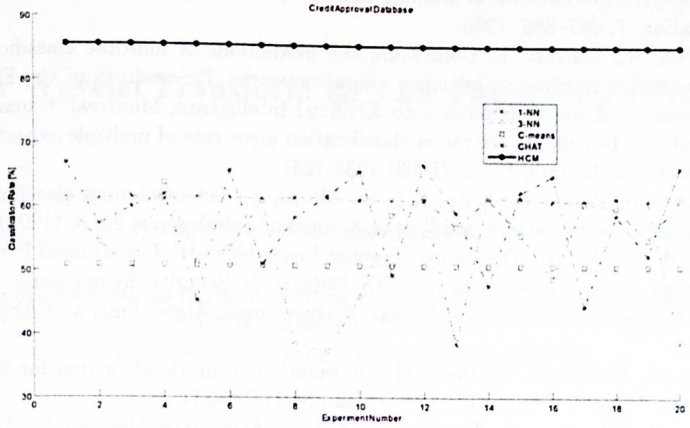


Fig. 2. Australian Credit Approval Database Classification Rate.

It is worth pointing out that this method does not have to compute a new classifier at each step, which represents an important reduction on computational costs. The learning phase of the process becomes swifter given the reduced number of input vectors considered. One and only one classifier is used throughout the entire process, which implies simplification on methodology. Classifier accuracy is not related to the number of input vectors considered in the learning phase, conversely, classifier accuracy is directly related to the mask value that allows elimination of redundant or irrelevant information.

Another clear advantage of this method is that the optimal mask search algorithm is applied only to those patterns that were previously considered during the learning phase, which means that no additional patterns are required to increase classifier accuracy.

This paper represents the initial works on optimized pattern classifier hardware implementation using feature selection and data dimensionality reduction.

Acknowledgments The authors of the present paper would like to thank the following institutions for their support: National Polytechnic Institute, Mexico (CIC, CGPI, PIFI, COFAA) , CONACyT and SNI.

References

[1] Nadler, M., Smith, E.: Pattern Recognition Engineering. John Wiley and Sons Inc., 1993.

[2] Nanni, L.: Cluster-based pattern discrimination: A novel technique for feature selection. Pattern Recognition Letters **27** (2006) 682-687

[3] Jacobs, R.: Methods for combining experts' probability assessments. Neural Computation, 7, 867–888, 1996.
[4] Maclin, R., Shavlik, J.: Combining the predictions of multiple classifiers: Using competitive learning to initialize neural networks. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada, 1995.
[5] Smyth, P.: Bounds on the mean classification error rate of multiple expert. Pattern Recognition Letters, 17, 12 (1996) 1253-1257
[6] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 3, (1998) 226-239
[7] Yan, W., Goebel, K.: Designing Classifier Ensembles with Constrained Performance Requirements. Proceedings of SPIE Defense & Security Symposium, Multisensor Multisource Information Fusion: Architectures, Algorithms, and Applications, 2004.
[8] Last, M., Kandel, A., Maimon, O.: Information-theoretic algorithm for feature selection. Pattern Recognition Letters **22** (2001) 799-811
[9] Palm, G., Schwenker, F., Sommer, F.: "Neural Associative Memories"., Associative Processing and Processors. Los Alamitos: IEEE Computer Society (1997) 307-326
[10] Dong, M., Kothari, R.: Feature subset selection using a new definition of classifiability. Pattern Recognition Letters **24** (2003) 1215-1225
[11] Gasca, E., Sánchez, J.S., Alonso, R.: Eliminating redundancy and irrelevance using a new MLP-based feature selection method. Pattern Recognition Letters **39** (2006) 313-315
[12] Kohonen, T.: Correlation Matrix Memories. IEEE Transactions on Computers. **21(4)**. (1972) 353-359
[13] Steinbuch, K. Die Lernmatrix. Kybernetik 1,1, (1961) 36-45.
[14] Anderson, J.A., Rosenfeld, E.: Neurocomputing: Fundations of Research, Cambridge: MIT Press. (1990)
[15] Hassoun, M. H. Fundamentals of Artificial Neural Networks, Cambridge: MIT Press. (1995)
[16] Santiago-Montero, R.: Hybrid Associative Classifier Based on Lernmatrix and Linear Associator (In Spanish). M.S. Thesis. Center for Computing Research, México (2003)
[17] Steinbuch, K. & Frank, H.: Nichtdigitale Lernmatrizen als Perzeptoren, Kybernetik, 1, 3, (1961) 117-124.
[18] Almuallim, H., & Dietterich, T.G. (1991).: Learning with many irrelevant features. Ninth National Conference on Artificial Intelligence (1991) 547-552. MIT Press.
[19] Shapira, Y., Gath, I.: Feature selection for multiple binary classification problems. Pattern Recognition Letters **20** (1999) 823-832

Haar Wavelet Transform for Low Power Video Frames

Héctor Silva-López[1] and Sergio Suárez Guerra [1]

[1] Center of Computational Research, National Technical Institute, México, A.P. 75-476, C.P. 07738, Zacatenco, DF, México.
esilvab05@sagitario.cic.ipn.mx, ssuarez@cic.ipn.mx

Abstract. Video compression is currently a prominent topic for both military and commercial researchers, due to the rapid proliferation of digital media and the subsequent need to store and transmit it in a space and time-effective manner. Most successful compression methods have been based on mathematically transforming an image (or sequence of images) into a frequency domain representation, and then filtering that representation to obtain a form suitable for effective encoding and compression. Our framework can be targeted to multimedia applications where video frames must be encoded, decoded and processed periodically within predefined video frame-rates. The frames arrive at the system dynamically, and leaves after a number of instances executed. The frames execute in a discrete voltage/frequency processor. This framework dynamically selects the sub-bands that maximize the energy savings in the system, such that no sub-band in the system misses its deadline. The problem is presented as a dynamic optimal problem with discrete constraints. Experimental results show that our algorithm present 29% energy saving.

1 Introduction

Although wavelet have their roots in approximation theory [7] and signal processing [14], they have recently been applied to many problems in computer graphics. These graphics applications include image editing [4], image compression [12], and image querying [3]. These applications impose strict quality of service requirements in the form of timing constraints. Ignoring energy consumption, operating the CPU at its highest speed operation quickly drains the batteries. Thus there is a tradeoff between reduced energy consumption and quality of service.

Voltage scaling technology has the potential to exploit such variability in the case of meeting timing constraints. By adjusting the operating voltage of the processor, the energy consumption and speed can be controlled [1]. Power regulator and variable voltage

© L.P. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 21-30

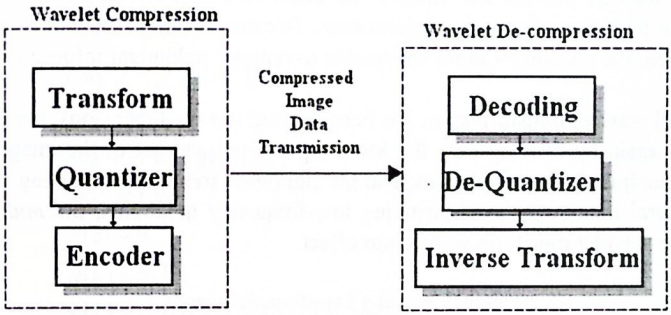
processors with response times in the microseconds range are available [10]. Fast response time makes it practical to dynamically adjust the voltage at run time.
Recently, researches have attempted to apply Dynamic Voltage Scaling (DVS) to video decoding to reduce power [5, 8, 9, 11]. These studies present approaches that predict the decoding time of incoming frames or Group of Pictures (GOPs), and reduce or increase the processor setting based on this prediction. As a result, idle processing time, which occurs when a specific frame decoding completes earlier than its playout time, is minimized. In [6] an alternative method called Dynamic is proposed as an improvement to these techniques. The Dynamic approach is designed to perform well even with high motion videos by dynamically adapting its prediction model based on the decoding experience of the particular video clip being played. The same authors present another alternative method called frame data computation aware (FDCA) in [2]. FDCA dynamically extracts useful frame characteristics while a frame is being decoded and uses this information to estimate the decoding time.
The objective of out work is to develop a novel DVS technique targeting such dynamic changing workloads. We presented a method that can be used to discern how much CPU throughout a wavelet transform is consuming. This information can be used to no deadline is missed and the energy saving is maximized for each sub-band when encoding a frame and decoding the same frame running on one processor.
This paper is structured as follows. In Section 2, we give a framework overview for the wavelets transform scheme. We then describe how is obtained energy saving for each sub-band, and the algorithm for selecting the optimal energy saving is showed in Section 3. Section 4 is an example showing how our scheme works on practical image. Section 5 presents the experimental results to demonstrate the performance of our low power video encoding/decoding scheme under different workload conditions. Section 6 summarizes our efforts.

2 Wavelet Image Compression overview.

Of the many processes available for image compression, two of the most popular transformations are the Discrete Cosine Transform (DCT) used in the common JPEG format, and the Discrete Wavelet Transform (DWT) used in the newer JPEG 2000 format. The DWT differs from the traditional DCT in several fundamental ways. The DCT operates by splitting the image into 8x8 blocks that are transformed independently [13]. Through this transformation process, the energy compaction property of the DCT ensures that the energy of the original data is concentrated in only a few of the transformed coefficients, which are used for further quantization and encoding [15]. It is the discarding of the lower-energy coefficients that result in image compression and the subsequent loss of image quality. Unfortunately, the rigid 8x8 block nature of the DCT makes it particularly susceptible to introducing compression artifacts (extraneous noise) around sharp edges in

an image. This is the "halo effect" seen in over compressed web images. Because the artifacts become more pronounced at higher compression ratios, JPEG's suitability for line drawings and cartoon-like images is significantly impaired.
In contrast to the DCT, the DWT operates over the entire image, eliminating artifacts like those caused by the 8x8 DCT blocking. Like the DCT, the fundamental wavelet transform is completely reversible, meaning that if the forward and reverse transforms are applied in sequence, the resulting data will be identical to the original. In addition, the DWT is based on sub-band coding where the image is analyzed and filtered to produce image components at different frequency sub-bands [18]. This produces significant energy compaction that is later exploited in the compression process. The wavelet's two-dimensional nature results in the image visually being divided into quarters with each pass through the wavelet transformation. A key effect of this transformation is that all of the high pass quadrants in the image contain essentially equivalent data [16]. A wavelet video compression, transmission, and decompression process that represents the target application of this vector processor is shown in Figure 1.

Figure 1: Wavelet Compression, Transmission, and Decompression Process



In this process, a single image or video frame is digitized by a camera and frame grabber. This image is then fed to a wavelet compression system. First, the compression system performs a wavelet transform of the image. This mathematical transform, which is perfectly reversible, converts the original image into a form suitable for encoding through a reversible spatial frequency separation. Next, the transformed image is quantized. This creates redundancy in the transformed image by reducing the number of allowable pixel values and thus the number of color or chrominance levels. Quantization is not reversible and is the principle cause of data loss in lossy compression. (The other cause of data loss can be attributed to floating-point rounding errors when calculating forward and inverse transformations with non-integer wavelets). The resulting bit stream from these two stages contains large blocks of redundant data that the encoder can easily locate and mathematically remove. At this point, a single image or video frame has been compressed into a bit stream that is some appreciable fraction of its original size, and is ready to be

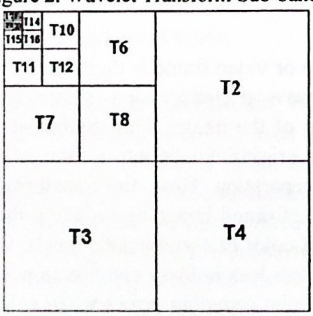
stored or transmitted to the destination system. Once at the destination, the process is reversed, whereby the image is first decoded, dequantized, and then inversed transformed to arrive at what is hopefully a convincing reproduction of the original image.

2.1) Wavelet transform.

In the field of wavelets, the modified Haar Wavelet (referred to as Haar*) is traditionally used for rudimentary image compression because of its algorithmic simplicity and low computational complexity due to an integer-based design [17]. When the wavelet is applied through its filtering process producing the scaling function coefficients (low-frequency) and wavelet coefficients (high-frequency). From the application of the Haar* wavelet, it is evident that the scaling function coefficient is simply the average of two consecutive pixel values, while the corresponding wavelet coefficient is the difference between the same two pixel values. The scaling function coefficients appear to contain all the image data, while the wavelet coefficients appear to be zero (black). If, however, the raw data was examined, it would be evident that the coefficients are only mostly zero. In reality, the wavelet coefficients contain the difference between adjacent pixel values, which is the high-frequency edge information. Because the high-frequency coefficients approach zero, the encoder is more easily able to remove redundant information from the image.

Once the full wavelet transformation has been applied to two dimensions, the process can be repeated again by filtering only the low-frequencies quadrant of the image. This low frequency quadrant is the visible image in the rightmost frame. By repeating the filtering process several times over ever-shrinking low-frequency quadrants, the *multiresolution* aspect of the wavelet transform comes into effect.

Figure 2. Wavelet Transform Sub-bands.



Because the multiresolution wavelet transform cycles by repeatedly processing the low-frequency information, some of the image data is processed more than once. Because of

this, the lower frequency wavelet coefficients are transformed by a wavelet of differing amplitude and duration than the higher frequency wavelet coefficients.
Thus, each sub-band shown in figure 2 was generated by a different wavelet function. It gives the different coefficient sub-bands. The Level 1 correspond T2, T3, and T4. Level 2 (T6, T7 and T8). Level 3 T10, T11, and T12. Level 4 (T14, T15, and T16) and Level 5 (T17, T18, T19, and T20).

3. Formulation of the problem.

The problem can be formulated as follows. Each time a new frame arrives at or leaves the system, the problem is to determine the mode (speed) of execution of the sub-bands such that no sub-band misses its deadline and the energy savings of the system is maximized. Each frame in the system execute in a discrete voltage/frequency processor. Note that a solution to this problem must be computed each time a new frame arrives or leaves the system, therefore a solution with probably cause deadlines to be missed.

The problem is formulated as:

$$\max_{\{u(k)\}_{k=0}^{2}} J = \sum_{k=0}^{2} S_k[u(k)]$$

subject to $x(k+1) = x(k) - u(k)$

$x(0) = 1.0,$

with : $x(3) = 0,$

$u(k) \le x(k),$

$u(k) \in \{1.0, 0.8, 0.6, 0.4, 0.15\}$ *for* $k = 0,1,2$

where

k = *correspond to sub – bands*

$S_k[u(k)]$ = *energy saving*

$x(k)$ = var *iable of state*

$u(k)$ = *control* var *iable,* (*speed to select*).

Bellman's optimality principle is used by compute the state variable in each stage, as follows:

Step 1: x(3) is calculated as follow:

$$J_3^*\{x(3)\} = 0$$

Step 2: x(2) $\in \{1.0, 0.8, 0.6, 0.4, 0.15\}$ is:

$$J_2^*\{x(2)\} = \max_{u(2)}\{S_2[u(2) + J_3^*\{x(2) - u(2)\}\},$$

where:

$$x(3) = 0 = x(2) - u(2), u(2) \leq x(2), u(2) \in \{0.15,0.4,0.6,0.8,1.0\}$$

Step 3: x(1) $\in \{1.0,0.8,0.6,0.4,0.15\}$, the equation that corresponds is:

$$J_1^*\{x(1)\} = \max_{u(1)}\{S_1[u(1) + J_2^*\{x(1) - u(1)\}\},$$

where:

$$u(1) \leq x(1), u(1) \in \{0.15,0.4,0.6,0.8,1.0\}$$

Step 4: x(0)=1.0 is:

$$J_0^*\{1.0\} = \max_{u(0)}\{S_0[u(0) + J_1^*\{1.0 - u(0)\}\},$$

where:

$$u(0) \leq 5, u(0) \in \{0.15,0.4,0.6,0.8,1.0\}$$

4. Example of the algorithm.

To illustrate the execution of the proposed algorithm, we consider the image Lena and we use five discrete speed levels {1.0, 0.8, 0.6, 0.4, 0.15}, in a discrete voltage/frequency processor. The CPU utilization can be measured while the system is under various states of loading. Obviously there is no way (yet) to measure CPU utilization directly. CPU utilization must be derived from measured changes in the period of the background loop. The average background loop period should be measured under various system loads and then the CPU utilization can be obtained. The CPU utilization is defined as the time 'not' spent executing the idle task (is a task with the absolute lowest priority in a multi-tasking system). The amount of time spent executing the idle task can be represented as a ratio of the period of the idle task in an unloaded CPU to the period of the idle task under some known load.

% time in the idle task = (average period of the background task with no load) * 100% / (average period of the background task with some load) [1]

Based on DVS technique, we adjust processor speed for each sub-band with slack reclamation.

In the table 1 present the time in micro seconds for every sub-band of the image. We consider 5 level how is presented in figure 2. The average background task execution time is 195 microseconds (because of space restrictions, we do not include here the details of how obtaining this time).

Table 1 shows too the results of applying equations 2 and 3 to the first and second columns in table 1.

Table 1. Sub-bands vs. Average background loop Period.

Sub-bands	T (microseconds)	% Idle	% CPU
1	974	20.02	79.97
2	535	36.49	63.55
3	388	50.26	49.74
4	859	22.70	77.30
5	417	46.76	53.24
6	353	55.24	44.76
7	595	32.77	67.23
8	368	52.99	47.01
9	287	67.94	32.06
10	445	43.82	56.18
11	313	62.30	37.70
12	267	73.03	26.96
13	342	57.02	42.98
14	281	69.39	30.61
15	247	78.95	21.05

After the executing the first sub-band of the level 1, to the maximum speed, we can obtain the percentage of utilization of this sub-band, the percentage idle is assigned to sub-bands 2 and 3, the executing speed the these sub-bands depending the result of first sub-band. Later, for the others levels, calculating the percentage of utilization for the first sub-band of each level and assigning the percentage idle to the two following sub-bands. In the table 2 is presented the percentage of energy saving level by level. The total energy saving of the all image is 44.19 %.

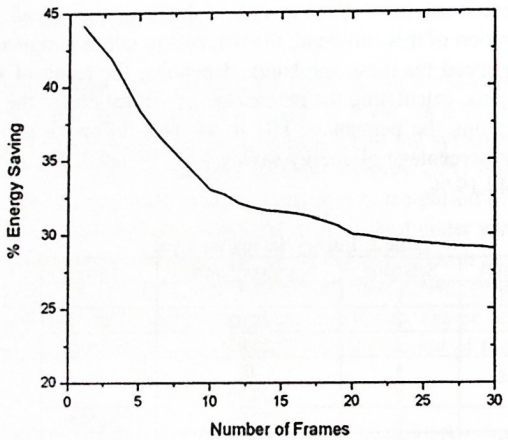
Table 2. Energy Saving for level.

Level	Sub-band	% Energy Saving	Total
1	1	0	
	2	20.02	
	3	20.02	= 37.50 %
2	4	0	
	5	22.70	
	6	22.70	= 36.40 %
3	7	0	
	8	32.77	
	9	32.77	
4	10	0	= 42.64 %
	11	43.82	
	12	43.82	
5	13	0	= 55.53 %
	14	57.02	
	15	57.02	

5. Simulation Experiments.

Our aim in this simulation experiments, was to verify the proposed algorithm in achieving our optimality criteria using different sizes of frames. The goals in this simulation experiments is to measure the energy saving over a large number of frames. Each value on figure 3 represents the results obtained from a set of 30 frames. We can observed that if the number of frames increase, the energy saving decrease until 29% of energy saving. On each frame generated, the metrics Energy Savings is obtained as a result of the execution of algorithm. The Execution Time metric denotes the execution time of each subband, which measures the physical time in microseconds, using a PC Intel PENTIUM Centrino 3.2 GHZ with 512MB of RAM and running on the Operating System SUSE Linux 9.3. The function used for the measurements is psched_get_time().

Figure 3. Energy vs. frames.



6. Conclusions

In this paper we proposed a power optimization method for a video encoder/decoder application running on a variable speed processor with five discrete speeds. The problem is presented as a linear problem with discrete constraints. Approximate solutions proposed are based on Bellman equation. Experimental results show that our algorithms yield near optimal performance with low complexity. For one frame is obtained 44.19% of energy saving but top 20 frames the energy saving is 29%. Also, as part of our future

work, we intend to extend our framework for running a real-time operating systems, improve the performance of our algorithm for increasing the energy saving of its performance and extend the framework to consider a large set of frames.

References

[1] A. Chandrakasan, S. Sheng and R. W. Brodersen, "Low-power CMOS digital design", IEEE Journal of Solid State Circuirs, 27, pp. 473-484, April 1992.

[2] B. Lee, E. Nurvitadhi, R. Dixit, C. Yu, and M. Kim, "Dynamic Voltage Scaling Techniques for power efficient video decoding", Journal of Systems Arquitecture, pp. 633-652, Available online 18 April 2005.

[3] Charles E. Jacobs, Adam Finkelstein, and David h. Salesin, "Fast multiresolution image querying", In Proceedings of SIGGRAPH 95, pages 277-286, ACM, New York 1995.

[4] Debora Berman, Jason Bartell, and David Salesin, "Multiresolution painting and compositing", In Proceedings of SIGGRAPH 94, pages 85-90, ACM, New York, 1994.

[5] D. Son, C. Yu, and H. Kim, "Dynamic Voltage Scaling on MPEG Decoding", International Conference of Parallel and Distributed Systems (ICPADS), June 2001.

[6] E. Nurvitadhi, B. Lee, C. Yu, and M. Kim, "A Comparative Study of Dynamic Voltage Scaling Techniques for Low-Power Video Decoding", International Conference on Embedded Systems and Applications (ESA), Jun 23-26, 2003.

[7] Ingrid Daubechies, "Orthonormal bases of compactly supported wavelets", Communications on Pure and Applied Mathematics, 41(7): 909-996, October 1988.

[8] J. Pouwelse, K. Langendoen, R. Lagendijk, and H. Sips, "Power-Aware Video Decoding", Picture Coding Symposium (PCS'01), Seoul, Korea, April 2001.

[9] J. Pouwelse, K. Langedoen, and H. Sips, "Dynamic Voltage Scaling on a Low-Power Microprocessor", 7th ACM Int. Confe, on Mobile Computing and Networking (Mobicom), pp. 251-259, Rome Italy, July 2001.

[10] M. Fleischmann, "Crusoe power management reduces the operating power with LongRun", in Hot Chips 12, Aug. 2000.

[11] M. Mesarina and Y. Turner, "Reduced Energy Decoding of MPEG Stream", ACM/SPIE Multimedia Computing and Networking 2002 (MMCN'02), San Jose CA, 18-25 January 2002.

[12] R. Devore, B. Jawerth, and B. Lucier, "Image compression Through wavelet transform coding", IEEE Transactions on Information Theory, 38(2):719-746, March 1992.

[13] Santa-Cruz, D, T. Ebrahimi, J. Askelöf, M. Larsson and C. A. Christopoulos, "JPEG 2000 Still Image Coding Versus Other Standards," *Proceedings of SPIE* 89 *45th Annual Meeting, Applications of Digital Image Processing XXIII*, Vol. 4115, San Diego, CA, July 30-August 4, 2000, pp. 446-454.

[14] Stephane Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7):674-693, July 1989.

[15] Subramanya, S.R., "Image Compression Techniques," *IEEE Potentials*, Vol. 20, No. 1, February/March 2001.

[16] Topiwala, P.N., Wavelet Image and Video Compression, Kluwer Academic Publishers, 1998.

[17] Villasenor, J., Belzer, B. and J. Liao, "Wavelet Filter Evaluation for Image Compression", *IEEE Transactions on Image Processing*, Vol. 4, No. 8, pp. 1053-1060, August 1995.

[18] Welstead, S, Fractal and Wavelet Image Compression Techniques, SPIE Optical Engineering Press, 1999.

Component for Debugging of Interrupt Based Systems

Alain Tamayo-Fong[1], Luis E. Leyva-del-Foyo[1&2], Pedro Mejia-Alvarez[2]

[1] Dpto. de Computación, Universidad de Oriente, *90500,* Santiago de Cuba, Cuba
[2] Dpto. de Computación, CINVESTAV-IPN, Av. I.P.N. 2508, México, D.F., 07300
alaint@csd.uo.edu.cu, leyvadelfoyo@yahoo.com, pmejia@cs.cinvestav.mx

Abstract. With the aim of developing and implementing a real-time operating system with experimental purpose, the development of a component that allows the debugging and verification of the correct behavior of the interrupt handling mechanism to its lowest level has been needed. This has been achieved through the simulation of interrupt requests at the hardware level. This module has been designed as an independent and reusable component that can be very useful not only for its original purpose, but also for the development of others embedded and real time systems. Besides, the component has a great educational value and as such, it has been used in laboratory practices in the teaching of the hardware interrupt mechanisms of PC systems.

1. Introduction

Hardware interrupts are an important part of computer architectures. Its purpose is to provide an efficient way for external devices to get the processor attention. Debugging interrupt-based systems is a vital subject due to of the inherent complexity of this kind of system. This complexity comes from the fact that interrupts may happen at any non predictable time, and can happen some of them at the same time.

Interrupt hardware assigns to each *Interrupt Request Line* (IRQ) a default static priority order. When an interrupt occurs, the task executed by the processor at that instant is interrupted. Execution control is transferred to the *Interrupt Service Routine* (ISR) assigned to the signaled IRQ line. While this routine is being executed, interrupt request lines with lower priority are disabled. An ISR can be interrupted only by a higher priority IRQ. Once served all IRQs, the interrupted task is resumed.

Computer systems using Intel processors and compliant with the industry standard [3], use an interrupt hardware compound by two *Programmable Interrupt Controller* (PIC) 8259 chips connected in cascade throw the IRQ2 line of the first 8259. This configuration provides 16 IRQ lines (IRQ0...IRQ15).

As part of the development of a *Real-Time Operating System* (RTOS) with special features concerning interrupt handling, the development of a component that allows for simulation of interrupt requests at the hardware level in a controlled way was needed. With this component we were able of debugging, testing, tuning and verification of the correct behavior of the interrupt handling mechanism. In spite of the existence of many low level debuggers [14], we did not find any debugging tool with the required features to accomplish this task.

© L.P. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 31-41

The main contribution of this work is the design of a module that allows the issue of hardware IRQs in any point of the code controlled by software. This avoids one of the greatest problems of interrupt-based systems: interrupt asynchrony. Furthermore, after the code is properly instrumented, the change of the sequence of interrupt request (test cases) is possible without further code modification. The module is an independent and reusable component that can be very useful not only to its intended purpose, but on the development process of others embedded and real-time systems.

The rest of this paper is organized as follow: Section 2 summarizes the novel features of the experimental RTOS that bring us to the need of this work; as well as, presents a brief survey of the related recent works. Section 3 explains the component design. Section 4 present a test case showing the way our component can be used. Finally, section 5 outlines the conclusion of our work.

2. Motivation for this work

This work is part of the development of the *Portable and Adaptable Real-Time and Embedded Microkernel Operating System* (PARTEMOS) [10]. This is an experimental microkernel being developed for experimenting novel interrupt and exception handling schemes for real-time and embedded systems [7, 9, 8].

2.1. Interrupt Handling in PARTEMOS

PARTEMOS presents some interesting special features, but for the purpose of this work we will only concentrate on those regarding its interrupt handling. The system transforms interrupts in signals operations over a semaphore [1], allowing the tasks to synchronize with interrupts through *wait*() operations on the semaphore associated with an IRQ line. In this way, PARTEMOS hides the (not desirable) interrupts asynchrony in the deepest part of the system. With this approach an IRQ can be seen as a "hardware task" that send a signal through its associated semaphore.

With the purpose of making the system portable, those aspects depending on the hardware have been located in the lowest layer named *Hardware Abstraction Layer (HAL)*. An important component of this HAL is the *INTerrupt HAL* (INTHAL) which isolates the rest of the kernel from the interrupt hardware, increasing the portability of the system. This layer presents the following specific features:

- Interrupts priorities do not map directly to default interrupt priorities assigned by the hardware. Hardware interrupt priorities can be rearranged as needed.
- Both interrupts and task are within the same priority space. Differently from most know systems where interrupts always have higher priorities than task, PARTEMOS allows tasks and interrupt priorities to be intermixed.

The implementation of these features has made the INTHAL a very complex layer, because the offered behavior is very far from that provided by the hardware. To achieve a correct implementation of this *virtual interrupt controller*, we need a component that eases its debugging. Specifically, we need a component that allows us to generate interrupts in a controlled way. In addition, this component must be flexible enough to build a wide set of test cases.

2.1. Related Works

Bug finding in interrupt driven code is one of the most complex and time consuming task of the overall development process of embedded systems. Recently, the general issue of making bug free interrupt driven code has been faced in different ways. In [15] a method for identify all possible sequences of execution of code in presence of interrupts is proposed, them standard sequential testing techniques can be used. A source code transformation technique presented in [12] turns interrupt-driven code into semantically equivalent thread-based code for checking with standard thread verifier. In [3] a technique for grouping code block for an effective use of model checking to analyze the behavior of interrupt-dependent programs was presented. The work in [4] presents an extension to Petri nets for modeling interrupts. Some techniques for debugging the interrupt driven embedded systems and for coping with the asynchronisms of interrupts were presented in [1]. Any of these techniques can not be used to test low level interrupt management or even the interrupt hardware itself. Existing development environment or debugging tools do not provide means for issue hardware interrupt over software control. The tools presented in this paper can issue hardware interrupt request under software control using a PC hardware trick similar to the used in at-hot way in [13] to test the interrupt behaviors in Windows NT. Our component can be used with any conventional debuggers for avoiding the asynchronisms in interrupt request, letting the repetition of test cases for easy finding of bugs under debugger control.

3. Component for the generation of hardware interrupts.

The designed component contains functions for interrupt generation on computer machines based on Intel Pentium processor or higher with a PCI chipset. Generation of hardware interrupts through software is implemented using an undocumented feature of these chipsets. Manipulating the interrupt hardware edge/level registers produce the activation of some IRQ lines [6]. At the beginning of the ISR serving interrupts generated in this way, the current state of these registers must be restored to its original state; otherwise, the interrupt signal will be generated continuously. It is important to emphasize that not all IRQ lines can be activated using this feature. Furthermore, the interrupts that can be activated this way may vary from machine to machine. For this reason, our component is divided in two parts. The first one is an executable program, named DOSCHK, whose purpose is to determine which interrupts can be generated on the specific machine. The second one is the library *IntrGen.lib* that must be linked with the program that needs to generate the hardware interrupts. This library provides mechanisms to use this feature in a convenient way.

The implementation of the DOSCHK program is simple, and consists in the manipulation of the interrupt hardware edge/level registers. Then, it must check which IRQ lines were activated. The library design is a more complicated issue. A convenient abstraction must be provided to library users, so they can build test cases in a simple and flexible way. This design is presented in the next subsection.

3.1. Library Design

The main objective of the library design is to provide users with a convenient abstraction that eases the building of test cases in a simple and flexible way. This objective includes several aspects as the introduction of minimal modifications to the debugged system, the possibility to change easily from a test case to another and the minimization of details that the person designing the test case must deal with.

A possible solution to this problem is to rely on a function that tries to activate a given IRQ line passed as parameter. This function must be inserted in key points of the debugged system. In this way we can determine how this system reacts to interrupts generated on these points. One advantage of this solution is that it is very easy to implement. Once selected the key points where interrupts must be generated, the interrupt sequence to be generated can be included directly on the source code or it can be stored on variables. The first choice causes that it is necessary to modify and recompile the system code for each test case, so this option is unacceptable. The second choice is more flexible because it allows users to change the sequence of generated interrupts changing the variables values. But, it does not allow users to change the number of interrupts generated on the selected points of the debugged system. The number of interrupts generated on each point must be known when the system is modified. This second choice is superior to the first one but is not flexible enough for our needs. We need to change not only which, but how many interrupts will be generated on each of these points, without changing the system code more than once.

The same problem arises when we need to check the state of the PIC. We could provide a function for returning its state. Calls to this function must be inserted at points of the debugged system where this information is needed. The problem results from the fact that this point might change from one test case to another. The choices are: modify the systems for each test case (which is not acceptable), or modify the systems just once and gather this information in all the places that might sometimes need it. Once again, the inflexibility of this solution is evident.

Another design solution is the definition of objects that encapsulate several interrupt generation requests. These objects must include a mechanism (methods) to be notified of which and how many interrupts must be generated in a given moment. This solution allows users to determine which points of the systems must be modified just once. Then, build a test case consist just of determining the values used to initialize this objects. Each object acts as a queue where interrupt generation requests are stored and provides methods to generate one or more interrupts.

To generate more than one interrupt at a given point we could rely on a method receiving as parameter how many interrupts must be generated from the queue. This solution is not flexible because it requires the number of interrupts to be specified on the source code directly or through variables. None of these choices is effective because one implies the modification of the system code for each test case, and the other, that we must control separately the interrupts to generate and the number of them, complicating the design of the test cases. The best solution is to rely on an integrated mechanism where an object could determine just from its state which and how many interrupts generate in each point. Following this reasoning we could include in the queue some marks to indicate how many interrupts must be generated

at each point. This solution simplifies the design of test cases because all the aspects regarding interrupt generation are part of a single object. In the same way, the request of information about the interrupt controller state can be included in this queue too. By including these requests on the queue object we gain total control about what action to execute at each important point of the system: generate a variable number of interrupts or obtain state information from the interrupt controller.

From this analysis we decide to implement a class representing a *command queue.* This queue will contain commands to generate interrupts, show state information of the interrupt controller and special markers to group commands so they could be executed in a single function call. This class will also contain methods to generate one or more interrupts. To generate a single interrupt a command is taken from the front of the queue and a function for activating the corresponding IRQ line is called. To generate more than one interrupt, commands are taken from the front of the queue and executed until a marker or the end of the queue is found.

3.2. Library interface

The *IntrGen.lib* exports the class *commmandQueue* declared in the header file *IntrGen.h.* This class implements the command queue using a structure containing an array of configurable size where commands are stored. The structure also contains the indexes of the first and last command on the queue. *commmandQueue* exports methods to: (1) Initialize the queue (*initCQ()*), (2) insert commands in the queue (*postCQ()*), (3) execute a single command from the queue (*execCQ()*), (4) execute several commands from the queue (*execAllCQ()*)

The initialization function *initCQ()* takes the maximum number of commands that the queue can hold and allocate the memory needed for holding the commands array.

The function *postCQ()* is used to insert commands in the queue. This function receives a variable number of parameters. The first parameter is the number of commands to insert and is followed by the command set.

The commands can be executed individually, calling the function *execCQ()*; or by groups, calling the function *execAllCQ()* that executes commands inside a loop. The command types that can be used are:

Activate an IRQ line
Request information from the PIC
Marker

The first command activates an IRQ line from the PIC; this is exactly what an external device does when it needs processor attention. These commands are named adding the interrupt line number to the identifier IRQ. For example: IRQ3, IRQ4, etc. The second command allows the users to read the PIC state. The 8259 PIC has three state registers, each one with a bit associated to each IRQ line: the *Interrupt Request Register (IRR)* indicates which IRQ lines are requesting the CPU attention; the *Interrupt Service Register (ISR)* shows which interrupts are being serviced; and the *Interrupt Mask Register (IMR)* indicate which interrupt lines are disabled. The PIC information can be read partially (one register at a time) using the CS_IRR, CS_ISR and CS_IMR commands; or totally using the CS_ALL command.

Markers provide a mechanism for grouping of commands that could be executed on a single function call. When the function *execAllCQ()* is invoked commands from the queue are executed until a marker or the end of the queue is found. If a marker is found when the function *execCQ()* is called, no action is taken.

Besides the commandQueue class, the library exports a function named *disableFLIRQ()*. This function must be called inside the ISR associated to interrupts generated through the edge/level register manipulation, for restoring these registers to its normal state. If this function were not called the IRQ line would continue generating this interrupt indefinitely.

4. Test Case using the Interrupt Generation Component

As an example, this section presents a test case for studying the interrupt controller behavior when nested interrupts occurs on the MS-DOS operating system. This operating system was chosen because it provides unlimited access to hardware. On the design of the test case, special care has been taken in avoiding the used of interrupts that might lead to system malfunctioning.

The first step is to determine which interrupts lines can be activated by software using the DOSCHK program. The generated interrupts might vary from machine to machine. A 166 MHz Pentium PC running MS-DOS 6.11 produced the output:

```
Checking interrupts activated through the edge/level registers...

activated IRQs: IRQ3 IRQ4 IRQ5 IRQ7 IRQ15
masked IRQs: IRQ3 IRQ4 IRQ5 IRQ7
```

This output shows that IRQ lines IRQ3, IRQ4, IRQ5, IRQ7 and IRQ15 can be activated by software. It also shows that IRQ lines IRQ3, IRQ4, IRQ5 and IRQ7 are masked. Next, we will use lines IRQ3, IRQ4 and IRQ5 to build an example test case.

4.1. Test Case Design

We will design a test case that shows the interrupt controller behavior when nested interrupts are generated using IRQ lines IRQ3, IRQ4 and IRQ5. This test case will illustrate what happens when an interrupt is being serviced and another one with higher priority is generated. It also illustrates what happens if a lower priority interrupt is generated. The test must reflect how the interrupt controller keeps track of the interrupts being serviced and the interrupt requesting the CPU attention.

The test case that will be implemented is the following: *Activate IRQ5; while its ISR is being executed, activate IRQ3. Finally, while the ISR that is serving IRQ3 is in action, activate IRQ4.*

The first thing to do is modifying the ISR associated to IRQ5 and IRQ3 because while they are being executed lines IRQ3 and IRQ4 must be activated respectively. The ISR associated to IRQ4 also needs to be modified to show interrupt controller information. The first interrupt is generated from the main program using the function *execAllCQ()*. This function is called inside the mentioned ISR to execute commands from the queue. Listing 1 shows how the modified ISR for IRQ3 looks like.

```
void interrupt irq3Handler()
{
   disableFLIRQ(3); // Reset edge/level register
   cprintf("Serving Interrupt 3...\r\n");
   execAllCQ(&queue);

   cprintf("Leaving Interrupt 3...\r\n");
   outportb(0x20,0x20); /* Send EOI*/
}
```

List. 1. Source code for a modified ISR.

The ISR calls the function *disableFLIRQ()* on its entry to reset the edge/level registers to its original state. This function takes as parameter the corresponding IRQ number. Then, the ISR tries to execute commands from the *commandQueue* object *queue* (declared previously in the program) using the function *execAllCQ()*. Once executed the commands the End Of Interrupt (EOI) signal must be sent to the interrupt controller. On entry and exit of the ISR, some messages are printed to help keep track of the execution flow.

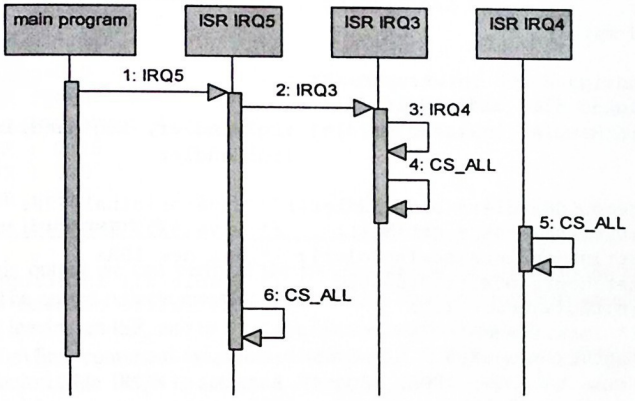


Fig. 1. Sequence diagram showing the expected execution flow. Arrows indicates execution control transfers when commands are executed. Markers are not shown in the diagram.

The hardest part of the process is choosing the right commands to achieve the expected interrupt sequence (Figure 1). The first action is activating line IRQ5, so this must the first command in the queue. This command is executed from the main() function. The next command activates IRQ3, in this way we can observe how the interrupt hardware behaves when a higher priority interrupt line is activated while a lower priority line is being served. In this case the ISR associated with IRQ3 must be executed immediately. The third command is IRQ4, so we can observe what happens when a lower priority interrupt line is activated while a higher priority one is being served. This request must wait until all interrupts with higher priority have been served. At this point the state of the interrupt controller is shown (CS_ALL), to take a look at the way the controller knows which lines are being served and which are requesting attention. The next command in the queue is a marker. This marker ends the execution of the function execAllCQ() inside the IRQ3 ISR. Once finished the

execution of this ISR the line IRQ4 must be served because its priority is higher than that of IRQ5. In this point we read the interrupt controller state once again (CS_ALL) and insert another marker to end up IRQ4 ISR execution. Then, we are back in IRQ5 ISR, now we can insert a marker to end up execution or just insert none because we are already at the end of the queue, so we will finish execution anyway. Once finished the IRQ5 ISR we return to the main() function.

4.2. Test Case Implementation

First, the source code of the program must be written (Listing 2). This source code must be compiled and linked with the *IntrGen.lib* library obtaining an executable file.

```
/* IntrTest.c: Test case for the interrupt hardware */
#include "IntrGen.h"
commandQueue queue;        /* command queue declaration*/
/* ISRs for IRQ3, IRQ4 and IRQ5 */
  ...          /* See Listing 1 */
void main()
{
   unsigned int interruptMask;
   intHandler oldIntHdler[3];
   intHandler newIntHdler[3]={ irq3Handler, irq4Handler,
                               irq5Handler               };

   readOldHandlers(oldIntHdler);/* Save original ISRs       */
   interruptMask = getIMR();    /* Save 8259 IMR register   */
   setIntHandlers(newIntHdler); /* Set new ISRs             */
   setInterruptMask(0x0000);    /* Enable all interrupts    */
   initCQ(&queue, 10);          /* Initialize command queue*/
   /* Insert command in the queue                          */
   postCQ( &queue,8    /* Number of commands               */
          ,IRQ5, IRQ3, IRQ4 /* Activate IRQ5, IRQ3 & IRQ4*/
          ,CS_ALL      /* Show 8259 state                  */
          ,C_BRK       /* Marker                           */
          ,CS_ALL      /* Show 8259 state                  */
          ,C_BRK       /* Marker                           */
          ,CS_ALL      /* Show 8259 state                  */ )
   execAllCQ(&queue);  /* Start commands execution  */
   setInterruptMask(interruptMask);/* Restore original IMR */
   setIntHandlers(oldIntHdler);   /* Restore original ISRs */
```

List. 2. Test case source code.

The program firstly includes the IntrGen.h header file. Next, a commandQueue object named *queue* is declared and the modified ISRs for the used interrupt lines are defined. On entry, *main*() save the addresses of the original ISR that will be replaced (IRQ3, IRQ4 e IRQ5) and the interrupt controller IMR register. These values are needed to restore the interrupt mechanism to its original state when the test is over. Then, we modify the *Interrupt Vector Table* (IVT) to point to the new ISRs, enable all

interrupt lines and construct the commandQueue object using the function *initCQ()*. At this point, we insert the commands for getting the desired sequence (*postCQ()*) and the execution is started calling the function *execAllCQ()*. Once finished the execution of commands we proceed to restore the IVT and the IMR register to its original state.

4.3. Test Case Execution Analysis.

The program execution produced the following output:

```
(0) Activating Interrupt 5
Serving Interrupt 5...
(1) Activating Interrupt 3
Serving Interrupt 3...
(2) Activating Interrupt 4
(3) ISR : 0x28
    IRR : 0x10
    IMR : 0x0
Leaving Interrupt 3...
Serving Interrupt 4...
(5) ISR : 0x30
    IRR : 0x0
    IMR : 0x0
Leaving Interrupt 4...
(7) ISR : 0x20
    IRR : 0x0
    IMR : 0x0
Leaving Interrupt 5...
```

From this output we can verify if the program execution adjusts to the expected behavior. The output reflects messages printed each time an interrupt is generated, on entering or leaving an ISR, and to show the interrupt controller state.

When the first command is executed (command 0) calling *execAllCQ()* at the *main()* function, line IRQ5 is activated. Then, its corresponding ISR takes control showing a message and calling the function *execAllCQ()* to activate the IRQ3 line (1). As we can see from the output, this interrupt is served immediately because of its higher priority. Inside the IRQ3 ISR line IRQ4 is activated (2), this interrupt is not served at this point because of its lower priority. This fact can be seen from the output. The next command executed while IRQ3 line is being served shows the interrupt controller state (3). The ISR register presents bits 3 and 5 set, reflecting that interrupt IRQ3 and IRQ5 are being served, which is correct. Register IRR presents bit 4 set, indicating that line IRQ4 is requesting the CPU attention. The IMR register value is not important in this test case; it keeps the same value all the time. After this, the next command, a marker (4), finish the execution of function *execAllCQ()* in IRQ3. The marker does not produce any output, but a message is printed when the ISR end up its execution.

The highest priority interrupt line waiting for attention is IRQ4, so its associated ISR is executed. This ISR executes the next command that shows the interrupt controller state once again (5). ISR register reflects that lines IRQ4 and IRQ5 are being served. The IRR register shows no line requesting attention. Right here, a

marker command ends up the command execution inside this ISR (6). At this point we are back to the ISR serving IRQ5. The next command to execute asks the interrupt controller for information about its state (7). We can see at the output that only IRQ5 is being served. There are no commands left in the queue so after showing a message execution control is back to the *main()* function. The program output proves that the interrupt hardware is working as expected when nested interrupts occur.

5. Conclusion

With this work we have obtained a very useful software component for the debugging and tuning of interrupt based systems. The main feature of this component is that it allows the generation of interrupts at the hardware level, but in a synchronous way under the software control. It is important to highlight that generated interrupts are not "simulated", but real hardware interrupts and go through all the PIC logic. As result, our component can be very useful for debugging and verification of the hardware interrupt handling mechanism itself, something that can not be done with the current tools. These tools use to "simulate" hardware interrupts using software interrupts and therefore, ignoring the interrupt hardware. As an example of the possibilities our component provides, impossible to achieve with other existent debugging modules, we can mention its utilization for getting the major temporal characteristics of a real-time system like the measurement of the interrupt latency without needing additional hardware. Additionally, the component has been used for teaching the interrupts topic at the computer architecture undergraduate courses on several computing majors.

References

1. Ball, S.: "*Embedded Microprocessor Systems*, Real World Design" Newnes 1998.
2. Dijkstra E. W., "*Cooperating Sequential Processes,*" Technical Report EWD-123, Technological University, Eindhoven, The Netherlands, (1965).
3. Fidge, C and Cook, P.:"*Model Checking Interrupt-Dependent Software*", In Proceedings 20th Asia-Pacific Software Engineering Conference (APSEC 2005), 2005.
4. Guangming, C., Minghong, L.: "*The Definition of an Interruptible Petri Nets*", Proc. of the 2nd International Conference on Embedded Software and Systems (ICESS'05), 2005.
5. Intel: "*IA-32 Intel Architecture: Software Developers Manual Vol. 3 System Programming Guide,*" 2002.
6. Intel: "*82371FB (PIIX) and 82371SB (PIIX3) PCI ISA IDE Accelerator*", Intel Developer's Insight CD-ROM, 1997.
7. Leyva-del-Foyo, L. E., Mejia-Alvarez, P., de Niz, D.: "*Real-Time Scheduling of Interrupt Request over Conventional PC Hardware*", Proc. of the Seven Mexican International Conference on Computer Science (ENC 2006), IEEE Computer Society 2006.
8. Leyva-del-Foyo, L. E.; Mejia-Alvarez, P.; de-Niz D.: "*Aligning Exception handling with design by contract in embedded real-time systems development*", Proceedings of the IEEE ECOOP 2005 Workshop on Exception Handling in Object Oriented Systems", 2005.
9. Leyva-del-Foyo, L. E, Mejia-Alvarez P.: "*Custom Interrupt Management for Real-Time and Embedded System Kernels,*" Proceedings of ERTSI: Embedded Real-Time Systems Implementation Workshop in the IEEE RTSS04, December 5-8, 2004 Lisbon, Portugal.

10. Leyva-del-Foyo, L. E.: "*Introducción a PARTEMOS.*" I Seminario Cientifico del Departamento de Computación. Universidad de Oriente, Julio de 2003.
11. McGivern, Joseph: "*Interrupt Driven PC System Design*", Annabooks/Rtc Books, 1998.
12. Regehr, John: "*Thread Verification vs. Interrupt Verification*", in Proceedings of the 2006 Federated Logic Conference (FLoC 2006), Seattle, August 10 - 22, 2006.
13. Roberts, D. : "*Interrupt Behavior in Windows NT*," Dr. Dobb's Journal, April 1998.
14. Rosenberg, J. B.: "*How Debuggers Work: Algorithms, Data Structures, and Architecture*", Wiley, 1996
15. Thane, H., Hansson H.: "*Handling Interrupt in Testing of Distributed Real-Time Systems*", 6th Int. Conf. on Real Time Computing Systems and Applications (RTCSA99), 1999.

Real-time Constraints for Adaptive Digital Filters on SISO Stochastic Systems

[1,2] P. Guevara López*, [1,3]J. J. Medel-Juárez+ and [2]B. E. López Silva*
(CICATA[1]-DEMS[2]-CIC[3]), Mexico Z.P. 07738,
*pguevara@ipn.mx, +jjmedelj@yahoo.com.mx. -*elizabeth_lopez_silva@hotmail.com

Abstract. A Real-time Adaptive Digital Filter (RTADF) interacts with dynamic process with non stationary output answer (in probability sense), having this filter two basic properties: response in quality [13] and response in time [24]: The filter has the critical dynamical restrictions knowing these as synchrony in each evolutionary interval (where the length of each interval is bounded), such that the time filter answer is bonded by each velocity moment, i.e., the filter adjust the time response with respect to the time evolution system, without lose time and information. In additional form but in the same way, the output filter answer has in each iteration an upper and lower limits: bounded these by measurable dynamical intervals [1], considering the properties exposed in [9], [10], [12], [13], [14] with respect to [6], [10], [22] and [27], such that the set of filter answers is bounded too with respect to Distribution Function predefined by the output system answers set, previously characterized. This basic property in probability sense is required for to construct a RTADF. Finally, this paper is organized considering the basic properties about the Real-time and Adaptive Digital Filter, exposing the basic ideas about the Real-time Adaptive Digital Filtering, developing these in the next section describing the RTADF in theoretical form with description in local and global constraints. The RTADF implementation: considering a D. C. motor, developing an extensive analysis of concurrent tasks: precedence constraints and times into a PC are considered as a part of the adaptive filter.

Keywords: *Adaptive Digital Filter, Synchronization, Real-time, Task, Interval.*

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 43-53

1 Introduction

The adaptive filter is required when neither the fixed specifications are known nor specifications can not be satisfied by time-invariant filters [10], [12], [21] [22] and [26]. Adaptive digital filter is a nonlinear digital filter since its characteristics are dependent on the input signal and consequently the homogeneity additively conditions are not satisfied [21], [22] and [26]. Generally, adaptive Digital Filters are applied in industrial processes into hardware that fulfillment the monitoring & control tasks (to see: [5], [7] & [12]). Adaptive Digital Filters have basic properties: response quality and b) response in time. In other case, the filter description or forecasting filter can be crashed in the real system where it interacts. These properties must be specified and implemented in a *Real-time Adaptive Digital Filters (RTADF)*. A RTADF can be implemented into embedded systems [4] and [12] using digital systems (micro controllers, DSP's, etc) and in digital computers with Real-time Operating Systems (RTOS) [18], [19] & [20].

2 Adaptive Digital Filters

Adaptive filter concept, is commonly used to describe a physic device characteristic, using the information given by the device (to see: [9] and [13]); such that, the filter is applied to a set of data generated by a corrupted signal (the desired signal is a "tricky task" and it is proposed, considering that it is bounded by the four first probability moments obtained by the corrupted signal, having both signal (corrupted and desired) the same distribution function). To talk about the information, it means to consider a lot of states of the system combined with a lot of perturbations (these perturbations have two sources around to the system: *a)* External and *b)* Internal. The noises or perturbations may arise from a variety of sources). The adaptive filter process is used on: a) Monitoring states (issued from the system), and b) Forecast or identify states without noises (originated from emitter, receptor or environment of the system). The adaptive filter theory has a basic area, which describe the main properties about the essential system characteristics (inside and outside qualities of it) [13], [21]: *Identification.* In [21] and [22], was considered two areas required for describing the inner dynamical (knowing as *Estimation*) and basic states; but in present time the estimate operation is used with optimal techniques to adjust the gains into identification filter, this step is knowing as feedback law (to see: [12] and [13]).

The adaptive filters use the information of the system; Error functional $J(k)_i$ (with respect to desired signal) must be converge to minimal value in finite time [6], [12], [13], [21] and [26]; the convergence criterion is defined by a difference between the filtering signal and desired signal and applying on this, the second probability moment [12], [13], [21], and [26], the criterion selected using the error functional, generally is part of the feedback law using to modify the gains into the basic filter or regular the input signal in the filter [7], [12], [13], [21], [22], and [26].

3 Real-time Digital Systems (RTDS)

A Real-time Digital Systems (RTDS) (to see: [4], [8], [11], [14], [20], [28]) could be described as *a digital system that suffice three conditions: a) interaction with physical world, b) correct responses, time constraints from physical world. RTDS* must be synchronized with physical world and this velocity is relative, depending of dynamics of real system and the adaptive filter algorithm, and the properties of the computing system. Then, *RTDS* may be fast or slow depending of real system dynamics; this obeys the criterions exposed in [24]. In a PC, the whole of all activities are processed by *Real-time tasks* [4], [18], [19], [20], [21].

4 Real-time Adaptive Digital Filters (RADF) - Introduction

In [7] was used a Real-time Adaptive Digital Filters implanted in *RTDS* considering the high velocity computers and facility to express the filter in recursive form. In [12] a FIR (*Finite Impulse Response*) Digital Filters was used as a basic "applications to Real-time processing of standard signals". In [25] that the implementation of Digital Filter into PC, was through to consider the times measure of all tasks around the implementation of it into PC (considering the measurable: The A/D and D/A converters, processor operations, filter algorithm and precedence constraints). But the concepts with respect to Real-time Adaptive Digital Filters weren't developed: Kalman Filter was studied in [18], but in it didn't justify the Real-time constraints.

5 Real-time Adaptive Digital Filter (RTADF)

Definition 1 (Real-Time Adaptive Digital Filter). A RTADF is an Adaptive Digital Filter with time constraints imposed by the dynamical process [24] and has the RTDF properties [23]. Including the main properties of it, the RTADF is described as:

Receiving and giving input and output responses, respectively, in synchronized form with respect to dynamics of the process. Inputs and outputs RTADF will be expressed in symbolic form as $\{u(k)\}_g$ and $\{y(k)\}_i$, with $g,i,k \in N^+$,[1[*]]. Giving correct responses set with respect to dynamical process: The quality response is defined in local and global senses considered in [12], [13], and [27] and must guarantee stable conditions defined in [2], [6], [9] and [13] and Express the RTADF in recursive form considering the concepts exposed in [3], [7] and [23], guarantee a minimal use of resources and memory, simulating the dynamics of the real process. The feedback law, bounded in time [24] and the quality of its answer [12], [13] and [22], observing the previous points described in this definition previously. The filter and the feedback law into the filter are described in convolution sense in discrete time and in recursive

---

* g, i, *represent a probably number of inputs and outputs respect in a concurrent system that evolution to k intervals.*

form, where the set of the changes and interactions between its tasks or instances are bounded, local and globally [12], [13], and [22].

6 RTADF's Constraints into a Digital Computer

When a RTADF in agreement to [23] is implemented in a computer with one processor, its different components use *concurrent Real-time tasks* (to see: [5] and [18]). The real-time task characteristics will be expressed in the following:

Definition 2 (RTDF: *local constraints*). The RTADF tasks have a lot of local constraints, imposed these by a lot of dynamical characteristics [11], [12], [18], [19], [20] and [24], with respect to real process:

a) Arrival time. ($I(k)_i$) [5]: Is the time when task becomes ready for execution. $I(k)_i > 0 \forall k \in N^+$., b) Computation time ($C(k)_i$): Is the total time in a instance (to see: [5]) of a task obtained by union of times used for atomic activities., c) Deadline minimum relative ($D(k)_{i_min}$): Is the minimum time in which a task should be finished[2]. It is a function of $I(k)_i$., d) Deadline Maximum relative ($D(k)_{i_max}$): Is the maximum time which a task should be accomplished[3], bounded it by the right side with sample time. This deadline is a function of time $I(k)_i$, and depend of the system evolution. e) Start time ($s(k)_i$): Is a time when the task starts with execution. This time depend of: Resource availability, latency times, context change, size of ready queue, etc. Start time is considered a stochastic variable [10] with a range named jitter,[18], and fulfill the condition $s(k)_i < D(k)_{i_max}$., f) Finishing Time (f(k)$_i$): Is the time when the task execution finished under interval time k. In mathematical sense $f(k)_i \in \left[ld(k)_{i_min}, LD(k)_{i_max}\right)$, with $ld(k)_{i_min} := I(k)_i + D(k)_{i_min}$, and $lD(k)_{i_max} := I(k)_i + D(k)_{i_max}$[4].,g) Lateness ($L(k)_i$): Is the loss time between finishing time and final interval time k. This time is $L(k)_i := \left| LD(k)_{i_max} - f(k)_i \right|$., h) Premature time ($P(k)_i$): Is the gain in time generated by finishes task before maximum deadline. This time is defined by $P(k)_i := \left| ld(k)_{i_min} - f(k)_i \right|$, where $f(k)_i \geq LD(k)_{i_min}$, generate a premature task answer., i) Sampling Period or Interaction time ($T(k)_i$): It is obtained by Sampling Criteria described in [24]. The RTADF evolution is bounded by Sample Period considering in each evolution the

[2] If the response is obtained before this lower deadline, it means that the system has a null task. In the other hand, if the time is higher that dynamic interval defined by dynamic system period time.

[3] If response is obtained after this deadline, it is bad.

[4] If we sum $I(k)_i$ to $D(k)_{i_min}$, $D(k)_{i_max}$, obtain absolute deadlines ($ld(k)_{i_min}$, $LD(k)_{i_max}$) [Liu00].

convolution between the filter properly speaking and the feedback law. The main characteristics of this are: **a.** $T(k)_t := 1/f(k)_{t_sampling}$. with $T(0)_t = T$, **b.** $\forall k \exists t(k)_t \cap t(k+1)_t - t(k)_t = T(k)_t$ and, **c.** $\mu\left[l(k)_t, LD(k)_{t_max}\right) := T(k) + \gamma_t$ where μ is a measurable function in the measure theory sense described in [1], [6] and [27], and γ_t, represent a tiny time (jitter). If $\left[l(k)_t, LD(k)_{t_max}\right)$ is empty, i.e., $\mu\left[l(k)_t, LD(k)_{t_max}\right) = 0$ at the sense described in [1].

Definition 3 (RTDF: global constraints**).** The RTADF whole tasks have a global deadline, considering that the infimum value in agreement criteria described in [12], [13], [19] and [22] could be closed with respect to supremum value allowed by dynamical process, and it is the equilibrium point, where the difference between the desired and the filtered signal tend to zero.

7 TRADF: Global performance

In this section will be described the RTADF global properties considering into the feedback law the convergence functional $\{J(m)\}$ when in it the infimum value tend to ε with $m > 0$. The number m represent the interval when the RTADF converge, and $(m_i \uparrow m$, with $i = \overline{1,n})$. The velocity converge depends the gains used into the feedback law considering that the values of $\{J(m)\}$ is lees than one.

Definition 4 (*Convergence: time* t_{i_c}). The time at which the RTADF converges, is $t_{i_c} := f_i(k = m)$, where m is number of the RTADF convergence interval and t_{i_c} has the condition $d_{i_c_min} \le t_{i_c} < d_i$.

When d_i is convergence deadline, where $d_{i_c_min}$ is defined as a minimal convergence deadline imposed by physical world. Guarantying with a feedback law a response on time and synchronized with physical world. The shortest minimal convergence time $(d_{i_c_min})$ is bounded with respect to the relative minimal deadline $(D(k)_{i_min})$, either.

THEOREM 1. The convergence function in probability sense [1], [2],[3], [6] and [27] defined by J_m in [6], [12], [13], [21], [22], [23] and [27], has a value ε_i semi-positive defined, with respect to convergence time described in symbolic form by t_{i_c}.

PROOF. By contradiction, suppose that ε_i is lower than zero $(\varepsilon_i < 0)$. The convergence error, defined in probability sense, is described by the second moment

(see for example: [1], [2],[3], [6], [12], [13], [21], [22], [23] and [27]) i.e., , $M\{\Delta_i \Delta_i^T\} \geq \rho_i$ with ρ_i positive semi defined, M represent the mathematical expectation operator and Δ_i is defined as difference between filtered value and real value (to see: [1] and [6]). Now, considering that the $\lim\left(M\{\Delta_i \Delta_i^T\}_{t \to t_i}\right) \to \varepsilon_i$, such that the superior limit of ρ_i when $t \to t_{i_c}$, is bounded by ε_i, and the inferior limit of ρ_i when $t \to t_{i_c}$, is bounded by zero. Then, $\varepsilon_i \geq 0$. □

8 RTADF Local behavior (for TVS)

All RTADF are stable (to see: [10], [13], and [15]) if the parameters gains are bounded into the disks [2] defined in the unitary circle for all k ([6], [17], and [27]): $\{a_e(k)\}_i \leq 1, e = \overline{1,n}$.

The optimal parameters obtained with respect the functional J_m and optimal criteria represent symbolically the proper values of modeled system [15] with desired digital signal answer, and these values are stable in discrete sense [2] and [6], without loss the disks into the unitary disk [6], [15], and [17]. Outside of this unitary disk the filter response is unstable, and the filter feedback law isn't convenient, [6] and [13].

THEOREM 2 (Relative maximal deadline $D(k)_{i_max}$). A RTADF fulfill $2f_{i_max}\left(D(k)_{i_max} - D(k)_{i_min}\right) < 1$, f_{i_max} is the dynamical process maximal frequency.

PROOF. Considering [24], the follows condition is true $f(k)_{i_nuestreo} \geq 2f(k)_{i_max}$, $2T(k)f(k)_{i_max} \leq 1$, where the $f(k)_{i_max}$ is the time the filter has convolution the filter with feedback law using in it the functional $J(k)_{i_max}$. And also, considering to Definition 2 section i, subsections a, b and c; the relative deadline differences is lower than sample time $T(k)_i > D(k)_{i_max} - D(k)_{i_min}$ and using transitivity in previous inequalities, we obtain the inequality $2f_{i_max}\left(D(k)_{i_max} - D(k)_{i_min}\right) < 1$. □

9 RTADF Computational times and Deadlines

A RTADF that is implanted in a digital computer with a one processor, the whole of the tasks around of filter will be scheduled in concurrent form; then the total time of RTADF will be conformed by the sum of computational times of total of tasks $C(k)_i = C(k)_X + C(k)_y + C(k)_{a \to X} + C(k)_J + C(k)_{au} + C(k)_{ay} + C(k)_{aye} + C(k)_{fb} + C(k)_{fb \to a}$, where: $C(k)_X$: Computation time of state equation algorithm., $C(k)_y$: Computation time of observable signal equation algorithm (desired signal)., $C(k)_{a \to X}$: Computation time of identification algorithm., $C(k)_J$: Computation time

of convergence error equation algorithm., $C(k)_{fb\to a}$: Feedback law, $C(k)_{au}$: Computation time of A/D conversion of input $u(t)$., $C(k)_{ay}$: Computation time of A/D conversion of output y(t)., $C(k)_{aye}$ C(k)$_{aye}$: Computation time of D/A conversion of the identification output.

THEOREM 3 (Computation time & deadline). A soft FDATR, the computational time $C(k)_i$ will be bounded by deadline $C(k)_i < D(k)_{i_max}$.

PROOF. a) Suppose the case with the following temporal constraints: $s(k)_i = l(k)_i$, $D(k)_{i_min} = 0$, $C(k)_i > D(k)_{i_max}$. $L(k)_i > 0$, $D(k)_{i_max} \le T(k)_i$. with the task time is described as (to see [20]): $C(k)_i = D(k)_{i_max} + L(k)_i$. Now, if $L(k)_i$ is expressed as submultiples of $D(k)_{i_max}$: $L(k)_i = D(k)_{i_max} r^{-1}$,where $r \in N^+, r > 1$.Then, the equality is: $C(k)_i = (1+\gamma)D(k)_{i_max}$, with $\gamma := r^{-1}$. The minimal value respect to relative deadline ($D(k)_{i_max}$), is $r = 1$, then this result implies that the starting hypothesis isn't comply. In accord to Definition 2, we have : $C(k)_i = f(k)_i - s(k)_i$, and $f(k)_i \in \left[l(k)_i + D(k)_{i_min}, l(k)_i + D(k)_{i_max}\right)$.Considering the worst case: $f(k)_i \to l(k)_i + D(k)_{i_max}$, and substituting in previous results we obtain: $l(k)_i - s(k)_i < 0$, such that this inequality is true, then the theorem is true. □

Corollary 1: In agreement to [19] about processor assignment, a hard RTADF is formed by a tasks whole such that finished before theirs local deadlines $\left\{D(k)^j_{i_max}\right\}$, and global deadline $D(k)_{i_max}$, where $D(k)^j_{i_max} \uparrow D(k)_{i_max}$, with $j = \overline{1,n}$.

Suppose that the global deadline obey the follows condition: $\exists i$ such as $C(k)_i > D(k)_{i_max}$. The assignment processor resources in accord to [19] are less to one, and then could be considered: $C(k)_i \left(D(k)_{i_max}\right)^{-1} \le 1$ this represents a basic contradiction because $C(k)_i < D(k)_{i_max}$ in agreement to Theorem 3, then this corollary is true.

THEOREM 4 (Synchronization). The RTADF whole responses would be synchronized with respect to temporal properties with respect to dynamical process, avoiding accumulative delay.

PROOF. The finished times set $\left\{f(k)_i\right\}$ would be bounded by a intervals set $\left\{ld(h)_{i_min}, LD(h)_{i_max}\right\}$ for all i,h and $k \in N^+$, such as for all $f(k)_i$, $f(k)_i \ge ld(h)_{i_min}$ and, $f(k)_i < LD(h)_{i_max}$. If and only if $h \equiv k$, and this equality implies that RTADF would be synchronized with dynamical properties of physical process.

By the other hand if: a) If $h > k$ then, the RTADF generate in it, accumulative delays and avoid synchronization. b) If $h < k$ then the RTADF generate accumulative advances and avoid synchronization. In the last two points, the RTADF tasks whole, don't have synchronization with physical process. The feedback law doesn't synchronized with the evolution of the input signal and state of the digital filter, then the results are bad because the answer of RTADF is obtained with delayed gains.□

10 RTADF Implementation Example in a Digital Computer

We will implement a RTADF using this with simple basic feedback law modifying inversely the gains of parameter evolution applied to D.C. motor.

To implement a RTADF into PC required a tool set:
a)D.C. Motor, 20 V, 1 A, 1800 rpm, two poles, permanent field., b) Power unit A/D, D/A, 5 V input, 20 V output, 0.05 A input, 5 A output., c) PC Pentium III 400 MHz, 64 MB RAM., d) ADC card PCL 818L, e) QNX® 4.24 Real-time Operating system., f) MicroPhoton Development Kit®.

11 RTADF Characteristics

The characteristics of model and RTADF are in agreement to [21]: a) SISO model (i=1), linear, nonstationary model in agreement to [2], [6], and [13],., b) Step input of 20 volts., c) Wiener Algorithm [27]., d) Noises v(k) and w(k) are correlated with observable signal described by y(k) but they aren't correlated between them., e) The dynamical model will be described as: $x(k+1) = a(k) * x(k) + \nu(k)$, $y(k) = x(k) + \omega(k)$, where: $x(k+1)$ is the system state, $y(k)$ is the observable signal (angular velocity), $\nu(k)$ and $\omega(k)$ are internal noise and external noise respectively and, "$a(k)$" is the parameter.

The identificator in basic sense is described as $\hat{y}(k) = \hat{a}(k) * y(k) + sing(\hat{y}(k) - y(k)) * J(k)$, Using the gradient as an optimal tool the estimator with feedback is expressed: $\hat{a}(k) := b(k)^{-1} p(k) + sing(\hat{a}(k) - a) * J(k)$, where, in recursive form , P(k)= P(k-1)+ y(k) y(k-1), B(k)= B(k-1) + y(k-1) y(k-1). The functional defined in [6] and [13] is expressed as: J(k)=E (Δ(k) Δ^T(k)), with $\Delta(k) := \hat{a}(k) - a$.

In the model described by (17) and (18), will be considering the next dates for it implementation [21]: a= 0.35, σ(v(k))= 0.95, σ(w(k))= 0.99.

12 RTADF implementation

The scheme tasks could be seeing in [4], [8] and [16]. Experimental implementation of FDTR is made by concurrent tasks: a) Maximal relative deadline $D(k)_{max}$, is equal to period T(k)., b) Bounded starting time (s_p-l(k) $\forall$k) was obtained by first probability moment, such as s_p = l(k)+0.0015 ms for each task., c) D.C. motor sampling period T(k) is 20 ms., d) Minimal relative deadline is $D(k)_{min}$= 2.5 ms., if the deadline is lower than 2.5 ms, the motor has bad operation., e) Convergence deadline is: d= 3 s. The RTADF require creating next task: Xk: System states algorithm, C_{Xk}= 0.237 ms, Yk: Observable signal algorithm, C_{Yk}= 0.289 ms, Bk: Observable signal variance, C_{Bk}= 0.268 ms, Pk: Ricatti equation algorithm, C_{Pk}= 0.249 ms, Ak: Parameter estimator algorithm, C_{Ak}= 3.252 ms, Jk: Error functional, C_{Jk}= 0.245 ms, FB: the feedback $C(k)_{fb\to a}$ =0.21ms, Au: A/D conversion of input, C_{Au}= 0.310 ms, Ay: A/D conversion of observable signal, C_{Ay}= 0.302 ms, Aye: D/A conversion of estimate signal, C_{Aye}= 0.314 ms, O: Parent task, C_O = 0.261 ms., C(k) is: 5.7305 ms. The computational times is described as $C(k)= C_{Xk} + C_{Yk} + C_{Bk} + C_{Pk} + C_{Ak} + C_{Jk} + C_{Au} + C_{Ay} + C_{Aye} + C(k)_{fb\to a}$,. And the times obtained: T(k)= 20 ms, l(k)= T(k-1)k ms, s(k)= l(k)+ 0.0135 ms, C(k)= 5.7305 ms, $D(k)_{min}$= 2.5 ms, $D(k)_{max}$= 20 ms, f(k)= 5.7305 ms, L(k)= 14.269 ms, P(k)= 3.2305 ms. The convergence time proved experimentally is: m= 113 intervals, t_c=2.45 s, d= 3 s, t_c<d is complied. The time value is 0.55 (To see: [21]).

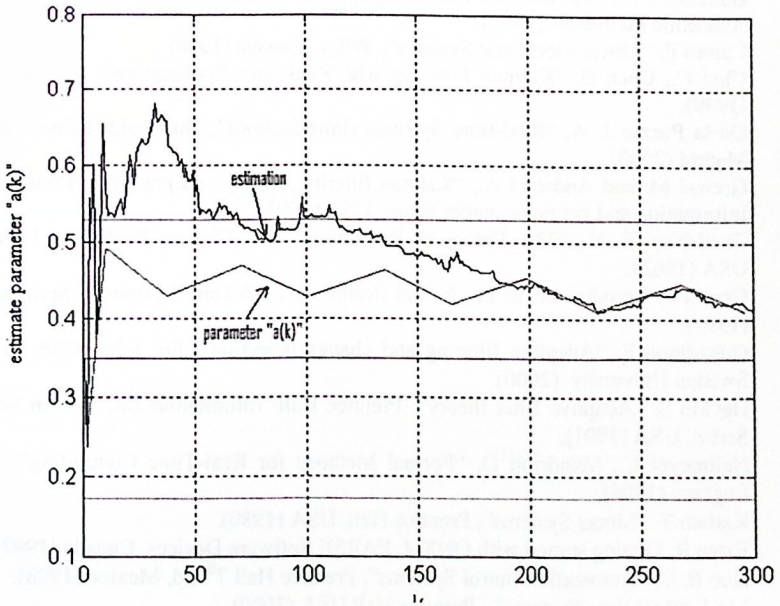


Fig. 1. RTADF: parameter estimator.

13 Conclusions

A Real-time Adaptive Digital Filter (RTADF) was described as a system that interacts with dynamic processes with non stationary output(s) (in probability sense), and has it two basic properties: response in quality [13] and in time [24]. The critical problem is the synchrony of it with respect to evolutionary answer into dynamic intervals.
In this paper we describe in theoretical sense the basic description about the RATDF, considering as a basic property the distribution Function about the output system answer, previously defined.
In practical sense we use the RTADF as estimator, and the illustration depicted, describe in illustrative form the convergence between the estimator and real parameter, with out the time properties that this filter require to accomplish.

References

[1] Ash R. "Real Analysis and Probability", Academic Press, USA (1970).
[2] Banach S., "The theory of Linear Operations", North-Holland Mathematical (1991).
[3] Baras J. "Symbolic and numeric Real-time signal processing", Technical Report University of Maryland, USA (1999).
[4] Burns A., Wellings A. "Real-time systems and programming languages". York University, Addison Wesley, UK (1997).
[5] Buttazzo G. "Hard real-time computing systems". Scuola Superiore S. Anna, Kluwer Academic Publishers (1997).
[6] Caines P. "Linear Stochastic Systems", Wiley, Canada (1986).
[7] Chui C., Chen G. "Kalman Filtering with Real-time Applications", Springer, USA (1999).
[8] De la Puente J. A. "Real-time Systems (Introduction)", Universidad Politécnica de Madrid (2000).
[9] Grewal M. and Andrews A., "Kalman filtering, theory and practice", Prentice Hall: Information and system sciences series, USA (1993).
[10] Gnedenko, B. V., "The Theory of Probability", Ed. Chelsea Publishing Company, USA (1962).
[11] Gray D. "Introduction to the formal design of Real-Time Systems", Springer, UK (1999).
[12] Gustafsson F. "Adaptive filtering and change detection". Ed. John Wiley & Sons, Sweden University (2000).
[13] Haykin S. "Adaptive filter theory". Prentice Hall: Information and System Sciences Series, USA (1991).
[14] Heitmeyer C., Mandrioli D. "Formal Methods for Real-Time Computing", Wiley, England (1996).
[15] Kailath T. "Linear Systems", Prentice Hall, USA (1980).
[16] Krten R. Getting started with QNX 4, PARSE Software Devices, Canada (1998).
[17] Kuo B. C. "Automatic Control Systems". Prentice Hall 7th Ed, Mexico (1996).
[18] Liu J. "Real-time Systems" , Prentice Hall USA (2000).
[19] Liu C., Layland J. "Scheduling algorithms for multiprogramming in hard-real-time environment". ACM- Journal. Vol. 20, No. 4 (1982) USA.
[20] Martín J. "Design Real-time computer systems", Diana, Mexico (1980).

[21] Medel J. J. "Análisis de dos métodos de estimación para sistemas lineales estacionarios e invariantes en el tiempo" Computación y sistemas volumen 6 número 1, México (2002).

[22] Medel J. J ., & Poznyak A. S., Adaptive Tracking for DC- derivate motor Based on Matriz Forgetting, C y S, (2001) 201-217.

[23] Medel J. J., Poznyak A. S., & Gevara L. P., Real-time Constraints for Digital filters on parameter estimate into stochastic systems, AVT, (2004).

[24] Nyquist H. Certain Topics in Telegraph Transmission Theory. AIEE Transactions, USA (1928).

[25] Papoulis A., Bertrán M. "Sistemas y circuitos", Ed. Marcombo, 1ª edición, España (1989).

[26] Paulo S. R. D., Adaptive Filtering: Algorithms & Practical Implementation, Kluwer Academic Publisher, 2nd Edition, (1997).

[27] Rudín W. "Análisis real y complejo". Universidad de Wisconsin, McGraw Hill (1988).

[28] Stankovic J. "Real-time Computing". Technical Report, University of Massachusetts, (1992).

Diagnostic of three-phase induction machine based on neural networks

Rodrigo López Cárdenas, Luis Pastor Sánchez Fernández, and Beatriz Adriana Jaime Fonseca

Center for Computing Research
National Polytechnic Institute, México
{rodrigo,bjaime}@sagitario.cic.ipn.mx
lsanchez@cic.ipn.mx

Abstract. Motors are essential components in most of today's manufacturing and production industries. Condition monitoring of electrical machines is a very important factor in achieving efficient operation of industrial processes. The rigorous requirements of modern electrical machines also necessitate condition monitoring systems, which enable the continuous monitoring of the system under all operations conditions. In this paper, characteristic patterns of well-operation three-phase squirrel cage induction machine are extracted from simulation, then varying some parameters, we can observe the behavior of induction machine to detect malfunctions.

1 Introduction

The induction machine is used in a wide of applications as means of converting electric power to mechanical work. It is without doubt the workhorse of the electric power industry. Pump and steel mill are applications of large multiphase induction motors, on a smaller scale, single-phase induction motors are widely used in household appliances as well as in hand and bench tools.

Three-phase squirrel cage induction motors are essential components in most of today's manufacturing and production industries. Safety, reliability, efficiency, and performance are some of the major concerns and needs for motor systems applications [1] [8].

Section two presents the three-phase induction machine mathematical model and justifies the mathematical transformation from *abc* variables to *dq* variables. Section three shows simulations of different ratings induction machine, section four deals with well-operation patters conformation and artificial network structure finally, section five is the conclusion.

2 Induction motor model

The idealized circuit model of a three-phase induction machine is described by six first-order differential equations, one for each winding, three of stator and

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 55-64

three of rotor. Although the rotor is a squirrel cage rotor, for simulation effects, is considered as winding rotor. These differential equations are coupled to one another through the mutual inductances between windings. In particular, the stator-to-rotor coupling terms are a function of rotor position, thus, when rotor rotates, these coupling terms vary with time [2].

Changes of variables are used in the analysis of ac machines to eliminate time-varying inductances, mathematical transformations as *dq* can facilitate the computation of transient solution of the induction model machine by transforming the differential equations with time-varying inductances to differential equations with constant inductances, moreover, this transformation reduces to four the number of electrical energy storage state variables, that is, only four first-order differential equations are needed instead six.

The transformation equation from *abc* variables to *dq* variables is given by

$$\begin{bmatrix} f_q \\ f_d \\ f_0 \end{bmatrix} = [\mathbf{T_{qd0}}(\theta)] \begin{bmatrix} f_a \\ f_b \\ f_c \end{bmatrix} \tag{1}$$

where the variable f can be phase voltages, currents, or flux linkages of the machine. The $dq0$ transformation matrix, $[\mathbf{T}_{qd0}(\theta)]$ is

$$[\mathbf{T}_{qd0}(\theta)] = \frac{2}{3} \begin{bmatrix} \cos\theta & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ \sin\theta & \sin\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{2}$$

and its inverse is

$$[\mathbf{T}_{qd0}(\theta)]^{-1} = \begin{bmatrix} \cos\theta & \sin\theta & 1 \\ \cos\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta - \frac{2\pi}{3}\right) & 1 \\ \cos\left(\theta + \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) & 1 \end{bmatrix} \tag{3}$$

One of the most popular induction motor model is Krause's model detailed in [3], the equations in flux linkage are as follows:

$$\dot{\psi}_{qs} = \omega_b \left[v_{qs} - \frac{\omega}{\omega_b}\psi_{ds} + \frac{r_s}{x_{ls}}(\psi_{mq} - \psi_{qs}) \right] \tag{4}$$

$$\dot{\psi}_{ds} = \omega_b \left[v_{ds} - \frac{\omega}{\omega_b}\psi_{qs} + \frac{r_s}{x_{ls}}(\psi_{md} - \psi_{ds}) \right] \tag{5}$$

$$\dot{\psi}_{qr} = \omega_b \left[v_{qr} - \frac{\omega - \omega_r}{\omega_b}\psi_{dr} + \frac{r_r}{x_{lr}}(\psi_{mq} - \psi_{qr}) \right] \tag{6}$$

$$\dot{\psi}_{dr} = \omega_b \left[v_{dr} - \frac{\omega - \omega_r}{\omega_b}\psi_{qr} + \frac{r_r}{x_{lr}}(\psi_{mq} - \psi_{dr}) \right] \tag{7}$$

$$\dot{\omega}_r = \left(\frac{p}{2J}\right)(T_e - T_L) \tag{8}$$

$$\psi_{mq} = x_M\left(\frac{\psi_{qs}}{x_{ls}} + \frac{\psi_{qr}}{x_{lr}}\right) \tag{9}$$

$$\psi_{md} = x_M\left(\frac{\psi_{ds}}{x_{ls}} + \frac{\psi_{dr}}{x_{lr}}\right) \tag{10}$$

$$\frac{1}{x_M} = \frac{1}{x_m} + \frac{1}{x_{ls}} + \frac{1}{x_{lr}} \tag{11}$$

where:

d : direct axis
q : quadrature axis
s : stator variable
r : rotor variable
ψ_{ij} : flux linkage
v_{qs}, v_{ds} : q and d -axis stator voltages
v_{qr}, v_{dr} : q and d -axis rotor voltages
ψ_{mq}, ψ_{md} : q and d magnetizing flux linkage
r_s, r_r : stator and rotor resistance
x_{ls}, x_{lr} : stator and rotor leakage reactance
P : number of poles
J : moment of inertia
T_e : electrical output torque
T_L : load torque
ω_e : stator angular electrical frequency
ω_b : motor angular electrical base frequency
ω_r : rotor angular electrical speed

For a squirrel cage induction motor as in the case of this paper, v_{qr} and v_{dr} are set to zero.

The induction machine dynamic equivalent circuit is shown in Fig 1, note that E_{qx} and E_{dx} are the unique electrical energy storage elements. The figure shows too the current direction and voltage bias of the transformed variables.

3 Model simulation

We start from the premise that we have only a mathematical model that supposes of well-operation induction motor. From this point, the simulations are done to various rating machine with free acceleration, that is, no-load. Then, varying some parameters, for example, resistance, we can observe the behavior of induction machine and extract particular conclusions.

The Table 1 shows various types of induction motor parameters, each machine is a 4-pole, 60Hz, 1800 rpm (synchronous speed) and three phase induction motor. The parameters are expressed in ohms using the 60Hz value of the reactances, the voltage is rated rms line-to-line voltage, and J includes inertia of the

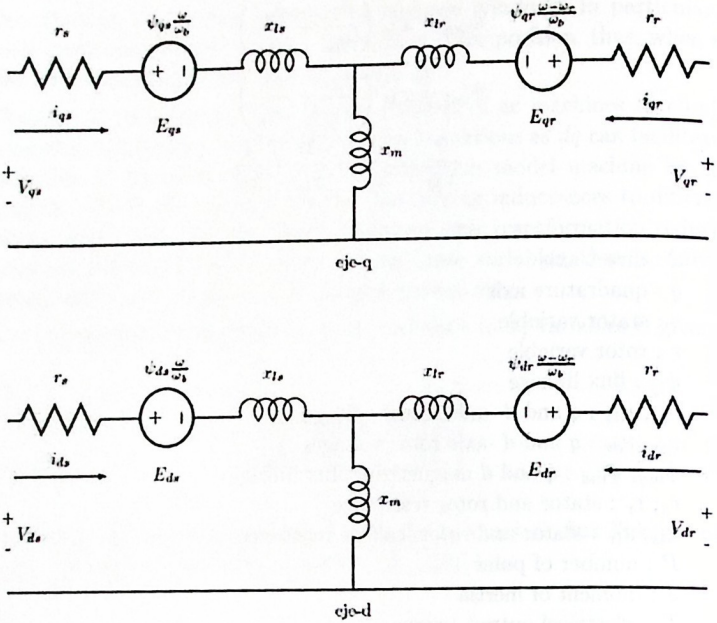


Fig. 1. Dynamic equivalent circuit of a induction machine

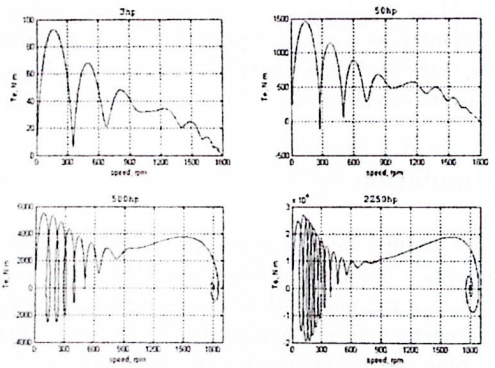


Fig. 2. Torque versus speed characteristics during free acceleration

load which is assumed to be equal to the inertia of the rotor. Rotor resistance and rotor leakage induction are referred to stator side [3] [4].

Table 1. Induction machine parameters

Motor Rating		r_s	r_r	x_m	x_{ls}	x_{lr}	J
HP	V	Ω	Ω	Ω	Ω	Ω	$kg-m^2$
3	220	0.435	0.816	26.13	0.754	0.754	0.089
50	460	0.087	0.228	13.08	0.302	0.302	1.662
500	2300	0.262	0.187	54.02	1.206	1.206	11.06
2250	2300	0.029	0.022	13.04	0.226	0.226	63.87

The differential equations that describes the induction machine were simulated on a computer using Matlab (like presented in [5]) with parameters of Table 1. The Fig. 2 shows the simulation reliability and depicts speed versus time characteristics during free acceleration (no load), Fig. 3 shows all machine variables of 3hp induction motor.

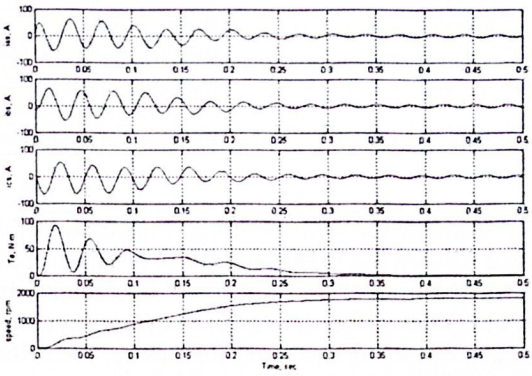


Fig. 3. Machine variables with no-load of 3hp induction machine

Because friction and windage losses are not taken into account, the machine accelerates to synchronous speed.

4 Characteristic patterns of operations

The characteristic patterns are conformed by nominal and dynamic variables. The nominal variables are manufacturer values, that is, horse power, line-to-line voltage, nominal current, revolutions per minute and number of poles; the dynamic variables can be extracted from simulation analysis or through physical measurements. Since non-invasively scheme proposed, we have suggested the

following signals and variables: peak current per phase (i_p), steady current per phase (i_n), time that each current reaches its steady value (t_n) and, time that rotor speed reaches its steady value (t_s).

Some malfunctions can be simulated varying machine parameters, for example, stator resistance affects motor efficiency, rotor resistance also affects motor efficiency but, had a profound effect on the shape torque-speed curve; increase the phase resistance of all coils simulates an increased temperature; inter-turn short circuit have a cumulative effect in decreasing the magneto motive force (MMF), as consequence the self-inductance and resistance of this phase winding changes, as does the mutual inductance between that phase and all others circuits in the machine; losses resulting from friction can be split up into Coulomb and Viscose friction, it can be simulated varying inertia parameter [7] [4] [6].

From above information, we can derive three kinds of situations: 1) motor efficiency, 2) winding performance and 3) friction problems. To simulate each of this situations we have modified certain parameters, the Table 2 show the condition types and the way of implementing.

Table 2. Condition types and its implementation

Condition	Condition Type	Simulation method
0	Normal	None
1	Motor efficiency	Increase rotor and stator resistance
2	Windings performance	Decrease stator resistance and stator leakage inductance
3	Friction problems	Increase value of inertia

The Fig. 4 - 6 confront the well-operating simulation (solid line) with malfunction simulation (dashed line). The patterns to recognize malfunction or well-operating are extracted from these results and are grouped in sets to train a neural network. The Table 4 show some results from simulation data without and with parameters variations such indicated above. From these data, we can observe that the principal alterations have effect on speed curve with a displacement from well-operating original curve and, with transient current, that is, increase o decrease of its peak value.

5 Malfunction identification using neural networks

For identification, an artificial neural network (ANN) with three layers is adequate due to their powerful nonlinear function approximation [8] [9] [10]. The input neurons are determined by number of nominal and dynamic variables, while number of output neurons are determined by the number of situations, that is, only three neurons. Since all motors are 60Hz, 4-pole and 1800rpm, the neuron inputs are: 1) horse power, 2) line-to-line voltage, 3-5) peak current for each phase, 6) steady state current and 7) rotor establishment time. The behavior of output neurons for each condition is shown in Table 3.

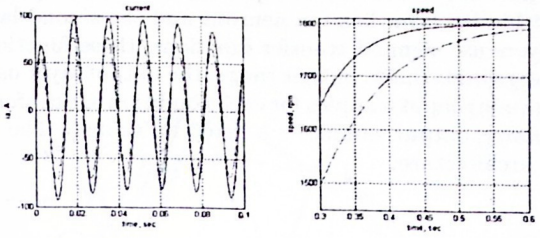


Fig. 4. Motor efficiency simulation for patterns extraction

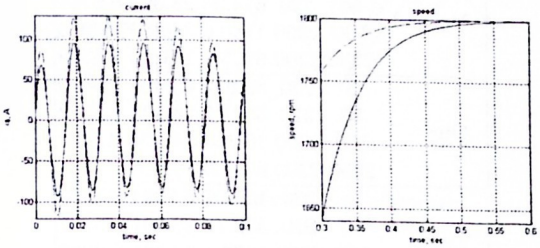


Fig. 5. Winding performance simulation for patters extraction

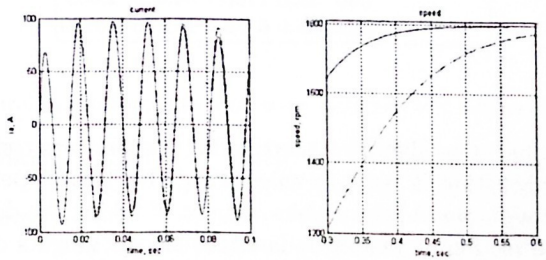


Fig. 6. Friction problems simulation for patterns extraction

Table 3. Training output samples

Condition	Neuron output		
	Neuron 1	Neuron 2	Neuron 3
Normal	0	0	0
One	1	0	0
Two	0	1	0
Three	0	0	1

The neural network consist of an input layer, output layer and two hidden layers. The first hidden layer has ten neurons and the second layer six neurons. Both hidden layers use sigmoid transfer functions, these functions compress an infinite input range into a finite output range, the output layer use linear transfer function. The training input samples for each condition are conformed as in Table 4, while the training output samples are listed in Table 3. The Fig. 7 show the neural network architecture.

Table 4. Patterns extracted from simulating data

Condition	hp	volt	i_{pb}	i_n	t_s
Normal	3	220	102.83	6.6943	0.3915
	50	220	674.20	28.097	0.5751
	500	2300	1160.6	34.023	1.4945
	2250	2300	6735.9	141.56	2.5369
One	3	220	80.037	6.6935	0.3415
	50	220	552.79	28.101	0.5312
	500	2300	1049.4	34.004	1.005
	2250	2300	6297.6	141.56	1.7035
Two	3	220	130.07	6.7904	0.3205
	50	220	552.79	28.101	0.5312
	500	2300	1572.4	34.381	0.9897
	2250	2300	9066.7	142.78	1.8453
Three	3	220	102.98	6.6946	0.585
	50	220	674.78	28.737	0.8531
	500	2300	1160.7	34.005	2.059
	2250	2300	6736.2	151.56	3.6299

The backpropagation algorithm used during training was resilient backpropagation. This algorithm is used to calculate derivatives of performance with respect to the weight and bias variables, consist of taking the derivative sign to determine the direction of weight update, the size of weights change is determined by a separately update value [11].

The Table 5 list the neural network output results. Of these data, several conclusions can be extracted: although all data are within a permissible rank, there are some deviation that neural network is difficult to identify. This drawback can be overcome increasing training patterns in number and description quality, in addition, increase the number of hidden layer can improve the results. Finally, the artificial neural network by itself can not provide heuristic knowledge of the motor because of its blackbox approach, therefore, a fuzzy logic system is a easily implement tool that utilizes heuristic reasoning and could provide extra solutions. Clearly, all patterns are not under noise because we take a snapshot of certain values, opposite case if we make an analysis of shape curve, where measurements can be affected by noise.

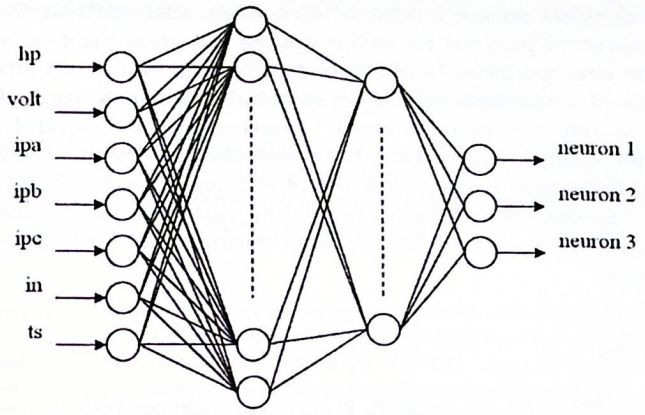


Fig. 7. Neural network architecture

Table 5. Neural network results with some patterns test versus expected values

Condition	3hp	50hp	500hp	2250
Normal	0.2539 — 0	0.3463 — 0	0.1133 — 0	0.1192 — 0
	0.2314 — 0	0.3401 — 0	0.1268 — 0	0.1258 — 0
	0.1773 — 0	0.0845 — 0	0.1159 — 0	0.1027 — 0
One	0.8252 — 1	0.9790 — 1	0.8802 — 1	0.9430 — 1
	-0.1540 — 0	-0.0539 — 0	-0.1070 — 0	-0.0539 — 0
	0.0529 — 0	-0.0257 — 0	-0.0719 — 0	-0.1099 — 0
Two	-0.1098 — 0	-0.0931 — 0	-0.0410 — 0	-0.0602 — 0
	0.8760 — 1	0.8681 — 1	0.9258 — 1	0.9431 — 1
	0.0233 — 0	0.162 — 0	-0.1034 — 0	-0.1029 — 0
Three	-0.1599 — 0	0.0740 — 0	-0.0462 — 0	-0.0230 — 0
	-0.1711 — 0	0.0827 — 0	-0.0530 — 0	-0.0274 — 0
	0.7969 — 1	1.1033 — 1	0.9683 — 1	0.9724 — 1

6 Conclusions

This paper described the construction of patterns to recognize different condition to which an induction machine can present. Through simulation analysis by means of mathematical models as Krause's model and transformations as *dq*, we constructed patterns for well-operating induction machine and possible malfunction circumstances. In order to identify between distinct situations, the applications of neural networks is an essential tool widely used, due to their capacities of approximation. In order to optimize results, several actions can be taken, for example, normalizing the values obtained by simulation can avoid the values' differences between low power and high power induction machine. Obviously, increase the amount of simulating patterns can accelerate the neural network convergence and finally, experimental results will suggest possible modifications.

References

1. Altug, S., Chow, M.-Y., Trussell, H.J.: Fuzzy inference system implemented on neural architectures for motor fault detection and diagnosis. IEEE Transactions on Industrial Electronics **46** (1999) 1069-1078.
2. Ong C-M.: Dynamic simulation of electric machinery. Prentice Hall PTR, USA 1998.
3. Krause, P.C., Wasynczuk O., Sudhoff S. D.: Analysis of electric machinery and drive systems. IEEE Press. Wiley-Interscience, USA 2002.
4. Armstrong.: Pump fault detection and diagnosis (FDD) based on electrical startup transient. Massachusetts Institute of Technology, Term Project, 2002.
5. Ozpineci B., Tolbert L. M.: Simulink implementation of induction machine model - A modular approach. IEEE International Electric Machine and Drives Conference, IEMDC'03. **2** (2003) 728-734.
6. Moseler O., Isermann R.: Application of model-based fault detection to a brushless DC motor. IEEE Transactions on Industrial Electronics **47** (2000) 1015-1020.
7. Joksimović G. M., Penman J.: The detection of inter-turn short circuits in the stator windings of operating motors. IEEE Transactions on Indutrial Electronics **47** (2000) 1078-1084.
8. Chow, M.-Y.: Methodologies of using neural networks and fuzzy logic technologies for motor incipient faults detection. World Scientific, Singapore 1997.
9. Quiang, S., Gao, X.Z., Zhuang, X.: State-of-the-art in soft computing-based motor fault diagnosis. Control Applications, 2003. CCA2003. Proceedings of 2003 IEEE Conference on Control Applications. **1** (2003) 1381-1386 vol. 2.
10. Filippeti, F., Franceschini, G., Tassoni, C., Vas, P.: Recent developments of induction motor drives fault diagnosis using AI techniques. IEEE Transactions on Industrial Electronics **47** (2000) 994-1004.
11. M. Riedmiller and H. Braun.: A direct adaptive method for faster backpropagation learning: The RPROP algorithm Proceedings of the IEEE International Conference on Neural Networks **1** (1993) 586-591.

Proposal for a Nanosatellite Microwave Communication System, to Achieve Energy Savings, High Capacity Data Transmission, and Position Control Feedback

Ante Salcedo[1] and Esaú Vicente-Vivas[2]

[1]Dpto. de Sistemas Digitales, ITAM, Río Hondo No. 1, 01000, México DF.,
asalcedog@itam.mx
[2]Instituto de Ingeniería, UNAM, Cd. Universitaria Coyoacan, 04510, México DF.,
evv@servidor.unam.mx

Abstract. This paper proposes the design of a microwave communication system for an under 10 Kgs Nanosatellite (NS) that is being designed and built at Instituto de Ingeniería - UNAM, Mexico. The proposed communication system was not contemplated in the original Nanosatellite proposal, nor in previous Mexican satellite projects, but it appears to be advantageous because of its narrow radiation pattern, and the high frequency at which it operates. As explained in the paper, we believe that the characteristics of such communications system will benefit the satellite with significant energy savings, high capacity for data transmission, and feedback information for position control. We present preliminary thoughts of the proposed communications system, describing the technology that is to be used for its implementation, and the benefits that such system brings to the nanosatellite. We also outline the next steps to be followed, so this paper should be considered as the starting point towards the design and implementation of the proposed nanosatellite microwave communication system.

Keywords: Nanosatellite, energy savings, position control, communications.

1 Introduction

Based on previous space project experiences such as the NASA´s Space Shuttle Get Away Special project developed by a Mexican team at Utah State University and the Satex Microsatellite project, the Instituto de Ingeniería (IdeI) UNAM took the initiative to develop an under 10 Kg nanosatellite with domestic technology, [1] and [2]. For these purposes IdeI has made informal academic alliances with several academic institutions from Mexico interested in developing satellite technology. With the Escuela de Aeronáutica IPN the structural design and the interface with the launching system are being worked. With the Centro de Investigación en Matemáticas (CIMAT) the satellite stabilization algorithms, the satellite dynamics model, the Earth magnetic field modelling, and a reference model that allows the satellite to be controlled from a

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 65-73

terrestrial geographic point are being elaborated. With the Instituto de Geografía UNAM and CIMAT a research laboratory is expected to be integrated to validate three axes satellite stabilization means and algorithms with instrumented frictionless equipment. With ITAM we agreed to develop directive communications systems to optimize following issues: employed energy, communications bandwidth and redundant means for satellite pointing control around the communication axis.

It is important to highlight that NS project aims the validation of subsystems and technologies that are fundamental for Mexico to develop bigger satellites in the medium term for Remote Sensing as well as for Telecommunications services. For those satellites the stabilization resources are essential either to acquire images of good resolution of geographic sites of interest, or, to efficiently intercommunicate geographic sites from our country and its continental region.

It should also be emphasized from one side that the Mexican satellite projects have been developed for Low Earth Orbit (LEO), and from the other that they have employed low data rate communications links for the purposes of satellite control and telemetry acquisition. Besides, LEO operation implicates a communications bottle neck because the satellites are continuously moving around the Earth. Furthermore they move to speeds faster than 28,000 Km/hour, whereas in the best of the cases the line of sight between satellite and Earth station is limited to 14 minutes. Particularly, the problem becomes augmented when satellites are required to download important amounts of data (images, special telemetry, etc). Under this scope the present paper proposes the development of Nanosatellite microwave communication system, for energy savings, high capacity data transmission, and position control feedback.

2 Efforts to design and build a Mexican Nanosatellite

There have been two important efforts towards the development, launching, and operation of Mexican experimental microsatellites: the UNAMSAT (~14 Kg) project [3], and the SATEX (~50Kg) project [4], [5], [6] and [7]. Both projects have been able to rise funding, generate Mexican expertise, produce a working prototype, and in the case of UNAMSAT it had a successful satellite launch. However, in both cases the projects came to an end, without continuing into further research and/or consecutive project developments.

In recent years, information technologies have evolved so much, that the capacities offered by microsatellites in the 90s can now be achieved with much smaller satellites (~5Kg), known as nanosatellites. This kind of satellites are encouraging for Mexican scientists, because they tend to be cost accessible (~60,000 $USD) and much easier to ensemble and deploy. Thus, in 2004, a new effort to develop a Mexican satellite (this time a 3.5 Kg nanosatellite) was born at IdeI, UNAM, [1] and [2].

The design and construction of the IdeI-UNAM Nanosatellite, expects to gather the expertise gained during the SATEX project, and to incorporate technology that is commercially available for the construction of nanosatellites. In a first step, the project team is looking forward to develop a proof of concept that includes as much as possible good quality parts and protections to ease the transition to a nanosatellite prototype and then gather credibility and support for the project.

2.1 The Mexican Nanosatellite

The nanosatellite envisioned by the IdeI-UNAM project team, is expected to be located at 750 Km from Earth, and to travel at such speed that it will circle the Earth 18 times a day. Different from geosynchronous satellites, the nanosatellite will pass above a ground station several times a day. When the satellite overpasses the ground station, it will only be in visible range during approximately 10 to 14 minutes.

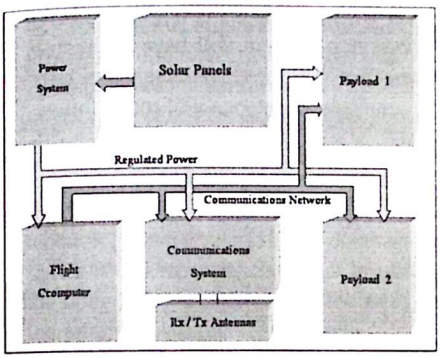


Fig. 1. Block diagram of NS subsystems

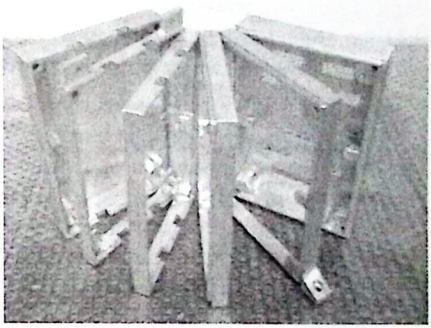


Fig. 2. Aluminium structural subsystem

Considering previous experience from the SATEX project, the proposed nanosatellite contemplates several subsystems, as illustrated in Figure 1. The payload systems are not necessary for satellite operation; and instead, they allow for experimental and test systems. While the payload systems have not been formally defined yet, there is a possibility that one of the payload modules may carry a position control system. Another payload module will be an infrared digital camera, so the satellite can broadcast aerial views.

Several subsystems of the satellite are in an advanced state of development, as they have been inherited from the SATEX project. Among these systems are the structural subsystem, flight computer, sensors, and software for the satellite operations and base station.

The structural subsystem holds together all other satellite subsystems, offering housing and protection to the circuits and systems in the satellite. The structure is composed by

several independent aluminium frames (modules) that hold and interconnect the different subsystems in the satellite, as they bunch together in a piggy-back fashion (see Figure 2).

The communications system chosen for the NS vehicle is inspired by the communication system used in the QuakeSat mission [8], and uses the TEKK KS-900/960 data radio and the Bay Pac 9600 modem. This communication system operates at 436.675 MHz (UHF frequency) for data transmission and reception; which (in the case of QuakeSat mission) allowed downloading 3 MBits per day. The radio has the following characteristics: 2 Watts RF output, varactor controlled Direct FM modulation, -30 to +60 °C operation, 0.199 Kg weight and 7.5-12 V operation.

While not completely defined yet, the Rx/Tx antennas chosen for the NS vehicle are short dipoles (1/4λ) made by spring steel tape measure material. The satellite will hold several of such antennas, to allow easy deployment and insure a broad radiation pattern for simple finding of ground station.

At the ground station, on the other hand, the proposed system will have a directive Yagi-Uda antenna with a position control system that will move the antenna to optimize the received power. In this fashion, the ground station antenna will rotate following the satellite (while it is in range of visibility).

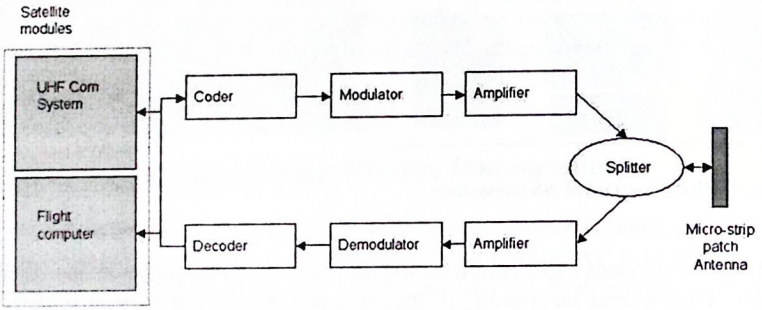


Fig. 3 Block diagram of NS subsystems

3 Proposal for a microwave communication module

As mentioned already, this paper proposes an alternative microwave communication system (MCS), based on high gain antennas operating at much higher frequencies (1-3 GHz). The proposed MCS is intended to become one of the payload modules that will allow communication with the ground station, providing a higher volume of data exchange, feedback for a satellite position control, and energy savings (if the main communications system is switched off after the MCS starts its operation).

To achieve an operating frequency between 1 and 3 GHz, the MCS will use micro-strip circuitry to feed a highly directive patch antenna that faces the ground station. If the patch antenna at the satellite can be accurately aligned with the ground station, the proposed MCS will constitute an excellent way to have efficient communications; however, small errors in alignment can reduce significantly the quality of communications.

Given that the satellite communicates with the ground station through the UHF system described in Section 1, the proposed microwave system won't be used until the satellite has been positioned to have one side facing Earth. Once the satellite is facing Earth, the MCS can be turned on, to achieve a quality of communications that surpasses the quality provided by the original UHF system (faster communications and more efficient use of available energy). In a first nanosatellite mission, the MCS can be loaded for experimental purposes only (as a payload); however, in consequent missions the UHF system could be used for initial communication with ground station, to achieve initial alignment before the microwave system is turned on and becomes the principal communication system (the UHF system can be turned off, or used for backup communications only).

Figure 3 shows the block diagram of the proposal. As shown, the system allows for transmission and reception through the same patch antenna. For the desired range of frequencies, the modules shown in the figure can be implemented with micro-strip technology, which tends to be compact (portable) and inexpensive. While the specific radios and components have not been chosen yet, they are commercially available and the complete system can be fitted within one of the aluminium modules considered in the satellite's structural subsystem.

4 Benefits for nanosatellite systems

As explained in the previous section, the major benefits that can be achieved with the MCS are: 1) the possibility to transmit higher amount of information, 2) the capacity to generate information to feedback the satellite position control, and 3) the possibility of saving energy, which in turn extends the satellites life or reduces the satellites weight. In this section, we further explain how to achieve such benefits.

4.1 High capacity communications channel

Like most nanosatellite projects, the nanosatellite being designed at IdeI-UNAM will carry a communication system that operates at 436.675 MHz (in the UHF band), and uses multiple linear dipole antennas to transmit/receive information. As explained before, this system is very convenient because dipoles have a broad radiation pattern that facilitates the establishment of a link between the satellite and the ground station

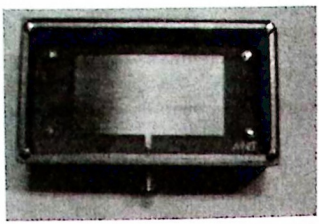


Fig. 4. Micro-strip patch antenna for 1-3 GHz [9]

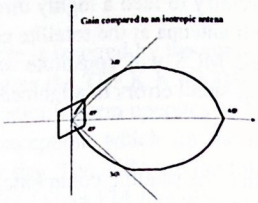


Fig. 5. Radiation pattern schematic for a patch antenna [9]

(i.e. the antennas don't have to be aligned). However, the frequency of operation is very low to establish a high speed digital communication link, and the broad radiation pattern is inconvenient because it wastes energy and is highly susceptible to noise.

With a carrier frequency of 436.675 MHz, the QuakeSat initiative at Stanford University, downloaded information at 9600 baud speed, which allowed downloading 3 MBits per day (~5 JPEG photographs, or one short video a day). This poor data transfer capacity will limit significantly the use of a digital camera (possible payload) to capture and download aerial views. Increasing the carrier frequency and the directivity of the communication antenna, allows to handle a larger communications bandwidth; which in turn allows a higher data transfer rate. With the 6dB gain characteristic of the patch antenna, we believe that the communications data rate could be at least doubled or tripled (6-9 Mbits/day). At the same time, as well as allowing a higher data transfer, the use of frequencies between 1 and 3 GHZ, allows the use of standardized wireless components, and a wireless digital communication protocol (such as WiMax).

The significant increment in data transfer capacity, joint with the possibility to use modern, standardized, commercial communication technology (components and protocols), makes very attractive the idea of using the proposed MCS and encourages the authors to further investigate and work towards its implementation.

4.2 Energy savings

As explained before, the use of dipole antennas is important to establish a communication link whatever their alignment is, because dipole based communication systems radiate energy into many directions (one of which needs to find the ground station despite the alignment of the dipole). However, since only the energy radiated towards the base station is utilized, the waste of energy is also very high. Dipole based communication systems use too much energy to achieve a specific signal to noise ratio (C/N) at the receiver location.

If instead of dipole antennas, the system used directive antennas (such as horns, apertures and/or micro-strip patches), most of the transmitted energy would reach the base station, thus reducing the amount of energy that needs to be invested to achieve a given C/N. The problem with directive antennas is that they need to be accurately aligned, or the quality of communication becomes poor and the energy waste becomes large.

Typically, while a linear dipole has a gain of ~0dB, a patch antenna has a gain of 6dB; which means that to have a given signal to noise ratio (C/N) the dipole needs four times more power than the patch. This is why, even if shortly after deployment the satellite requires a dipole based communications system, in the long run it is desirable that the satellite communicates using directive antennas, because the enormous power/energy reduction factor (1/4) result in large energy savings, which have a direct impact on the expected satellites life.

Again, the quick calculation of energy savings encourages further investigation, and supports the idea that the MCS is a good option for an experimental purposes payload.

4.3 Feedback for position control system

Finally, the narrow radiation pattern that characterizes the patch antenna, and that makes difficult the initial satellite positioning, turns out to be a very desirable quality for control purposes. First of all, as explained before, the high directivity of the antenna reduces significantly the energy waste during transmission; but also the linear decay of the antenna gain due to small variations from the optimal alignment, produces a decay in the transmitted/received power, thus giving valuable information to feedback the satellite position control system.

In order to have a stable feedback for the position control system, the ground station is being designed in such fashion that it will produce a uniform (omni directional) radiation in the direction of the satellite; in order to have the satellite control system forcing the satellite to face ground station, rather than having (as current nanosatellite projects) a directional antenna on the ground station with a control system that makes it face to the satellite. In principle, the ground station will have a spherical dome with micro strip patches all over it (which will be presented in future works), just for control purposes.

Considering that typical radiation pattern for patch antennas exhibits an approximately linear response to small alignment variations (see Figure 5), and that the intensity of received power (Watts) can be easily measured and sampled, the received power constitutes an excellent variable to be monitored and used to feedback the position control system.

5 Discussion and conclusions

While this paper proposes to incorporate a microwave communication system to the IdeI-UNAM nanosatellite (which was not contemplated before) and highlights its benefits, the task to do is not easy. There are indeed many problems to be solved, and calculations to be done before a clear conceptual proof of concept design is completed.

Among the many tasks that need to be carried out, we need to come up with accurate calculations of communication links, we need to evaluate propagation effects (such as attenuation) at the operating range of frequencies, we need to come up with a better estimation of the noise levels and an accurate calculation of the power, signal to noise ratio, bandwidth and data transfer rates. We need to build a model to understand variations due to distance and alignment errors, and to understand how the misalignment information can be used to correct position. We also need to establish commercial components that can be used to implement the system and carry on short experiments and tests to demonstrate that the proposal can bring the promised benefits. On the other hand, even if not explained in this project, we need to work on a detailed specification of the ground station communications base, which is crucial to achieve stable control feedback, and a significant bandwidth increase.

The work ahead is long and hard; however, we know that the technology for implementation is available, and the quick calculations summarized along this paper make us believe that the proposed system will contribute the IdeI satellite with: good energy savings, high capacity communications channel, and valuable position control feedback. We have no doubt that the proposed system is viable, and that the work necessary is worth, as the prize at the end of the road is sweet and encouraging.

References

1. Esaú Vicente-Vivas etal, "Preliminary Design of the Power Subsystem for a University Nanosatellite", CICINDI 2005, **Revista: Research On Computer Science,** Advances on Digital System Design, Vol.14, pp. 116-126, ISBN: 970-36-0180-4, ISSN: 1665-9899, Centro de Investigación en Computación, IPN, México, DF, **September 2005.**
2. Esau Vicente Vivas etal, "Design of a Nanosatellite Laboratory Model as a Proof of Concept for a future University Space Mission", **Revista: Research On Computer Science,** Advances on Digital System Design, Vol.14, pp. 105-115, ISBN: 970-36-0180-4, ISSN: 1665-9899, Centro de Investigación en Computación, IPN, México, DF, **September 2005.**
3. Unamsat, http://space.skyrocket.de/doc_sdat/unamsat-a.htm, (1996).
4. Satex, http://pumas.iingen.unam.mx/satex/, (2003).
5. Vicente-Vivas E. et al., *"La Convergencia de un proyecto multi-institucional para el desarrollo de un Microsatélite con*

tecnología Mexicana", Revista INGENIERÍA Investigación y Tecnología, ISSN 1405-7743, pp. 157-168, Vol III, No.
4, Octubre-Diciembre (2002).

6. E. Vicente-Vivas y D.F. García Nocetti, "Computación Tolerante a Fallas Aplicada en el Mantenimiento Automático del Microsatélite Satex", **Revista Iberoamericana Información Tecnológica, ISSN 0716-8756, Vol. 17, No 1, pp. 61-68, Enero 2006, Chile.**
7. Esaú Vicente-Vivas, Fabián García-Nocetti and Francisco Mendieta-Jiménez, "Automatic maintenance payload on board of a Mexican LEO microsatellite", ***Acta Astronautica Journal, Elsevier Science,*** *Volume 58, Issue 3,* ***February 2006,*** *Pages 149-167.*
8. Long M., et al., *"A CubeSat Derived Design for a Unique Academic Research Mission in Earthquake Signature Detec*
tion", Proceedings of the 16th Annual/USU Conference on Small Satellites, Logan, Utah, August (2002).
9. Mcirostrip Trainer (MST532) handbook; Feedback Telecommunications; E. Sussex; England.

A Method to Classify Digital Clocks Employing Image Segmentation and Associative Memories

José Luis Oropeza Rodríguez[1]
Sergio Suárez Guerra [2]

[1] Center of Innovation and Technological Development in Computation
Av. Te ·950, Edificio de Graduados 2° piso.,
Col. Granjas Mexico, Mexico, D. F.,
jloropeza@ipn.mx
[2] Computing Research Center
Av. Juan de Dios Batiz esq. Miguel Othon de Mendizabal s/n
Col. Nueva Industrial Vallejo, Mexico, D. F.
ssuarez@cic.ipn.mx

Abstract. This paper shows results obtained in the classification task of digital clocks in a real experiment for the recognition of digits extracted from a sequence of images taken from an intelligent digital camera (VCM40), mounted on a pneumatic machine. The purpose of the experiment is to obtain a smaller number of rejected pieces.

The digital clocks have an own sequence themselves, who depends of the type of clock that is being proven. For that reason, for each digital clock, the system has to classify them at the same time to check them and obtain a result of their sequence to verify that its working good or not.

Then, the problem to solve is simply to try to identify what kind of clock is mounted on the pneumatic machine, proving the sequence of each clock and to proceed to reject it or not. The neural network is employed to realize the classification task of the digits showed during the states of the sequence and to be compared with a conventional method used in the industry.

The results obtained were satisfactory because of the number of rejected pieces was adequately selected. The new method in comparison with the method used in the last process implemented, demonstrate that the Neural Networks application is very interesting for this industrial process. While the last method rejected 10% of which 4% was perfectly, the new method only have an error rate of 5%.

1 Introduction

Immersion of a visual machine to agile the industrial processes is a consequence of the utility of the computer science in our days. The faculty of the computer to see, to

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 75-84

hear and to move it is a reality in our days. Also, robotics industry is growing day to day and its contributions too. In this paper we are going to describe a process to classify a set of digital clocks manufactured by ELTEC Company. The process to check the different types of digital clock depends of the ELTEC requirements. In this analysis we used the digital clock named TOD-2 M.

At the same time, this paper shows the interactive image recognition in an industrial process that counts with a set of sequences predetermined conforms by 9 stages. The image recognition process was made in various stages of them (6 in total). The objective is that the sequence must be adequately finished detecting any problems while is making. When a problem has been found is registered in a database, so the system has a register of the all problems found during the process.

Immersion of the vision computer in an industrial process, give us the following advantages:

a) The number of pieces analyzed is greater than if a human supervisor making the same task.
b) The interested people have a register adequate of the problems founded in the electronic board analysis. With this, they must try to obtain the cause that origin the problem and then to try to solve it.
c) To implement different types of digital signal processing techniques to the industrial applications.

As we mentioned above and in [1], we try to apply the vision computer in a real world application and obtain the best response of the equipment.

The classification of segmented image regions for object recognition has been addressed in several works. Neural Networks are used in [2] not only to classify known objects but to detect new image object. In [3], object of interest are first localized, the feature is extracted from the region of the interest and finally a Neural Network is applied to classify objects. Support Vector Machines are used in [4] to classify a segmented image region in two categories.

2 Characteristics and Generalities

Table 1 resumes the characteristics of the TOD-2 M, clock used during this experiment and figures 1 shows digital clock TOD-2 M.

TIMER	• It shows the time of the timer if this one previously has been programmed. • If then it is marking and showing the time programmed in the timer (TIMER), the way of programming of the same one activates and it reestablishes it to zero minutes (00) • If one is in the way of programming of the timer, reestablishes the time to 00 and it remains within the same way.
▲	• When pressing momentarily this button, the time is increased by one minute. • If this button stays pressed by more than 1 second, the time is decremented in intervals of 10 minutes. • When the indicator of minutes reaches a count of 60, the hours will be increased automatically and the minutes will be reestablished to 00. • When is being the time of the timer, it allows to make changes to the value of the same one in real time.
▼	• When pressing momentarily this button, the time decrement by a minute. • If this button stays pressed by more than 1 second, the time is decremented in intervals of 10 minutes. • When the indicator of minutes reaches a count underneath 60, the hours will be to decrement automatically and the minutes will be reestablished to 59. • When is being the time of the timer, it allows to make changes to the value of the same one in real time.
CLOCK	• It activates the programming of the clock or returns to show the present hour if the TIMER is activated. • It accepts the new hour programmed when it is in the mode of programming of the clock. • It can be used like the key of beginning for any mode of programming.

Table 1 Operation each button inside the digital clock TOD-2 M.

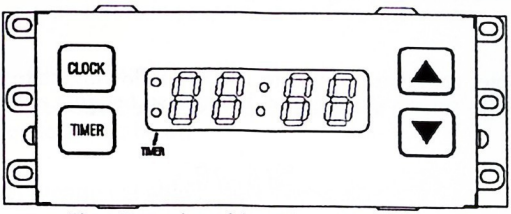


Fig. 1 Front view of the digital clock TOD-2 M

A simplified block diagram of the system is shown in figure 2.

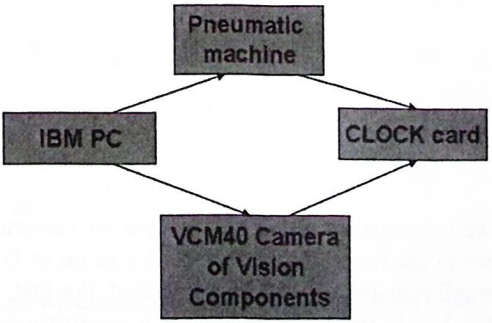


Fig. 2 Block diagram of the purpose system.

The VCM40 is manufactured by Vision Camera from Germany. It is an intelligent camera and it has three electronic boards into it. One of them has the DSP that controls all process concerning to the digital image processing task; it uses a specific program to interact with external equipment and it also counts with an operating system owned. The operating system has a set of instructions relatively small (it has only 40 instructions to interact with the user). An interface of communications is used to send the commands to the camera. The software included has the PROCOMM program that has the same utility of Windows HyperTerminal program.

Also, the user can interact with the camera using a C program compiled with the g21 compiler. This compiler was made for Analog Devices to interact with its commercial products. The experiments were executed on Windows 98 because of the compiler does not support the NTFS architecture.

The limitation of this camera is that only can take 14 pictures because DRAM is limited. Then, the problem is to try to adapt the number of images that we must take with the purpose to have good performance of the different digital clocks that are proved.

We prefer to solve this problem, using the sequence and use cases diagrams to relate the real problem with its solution. As we said, the camera only can be programmed using C language, for that, we do not use all types of diagrams that UML uses to represent the software engineering solutions. For that, we decided use only two of them.

At the same time, we have decided to use the following state sequence to represents the states of the TOD-2 M digital clock, as figure 3 shows.

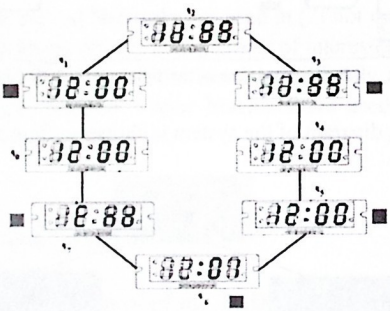


Fig. 3 State sequences for digital clock TOD-2 M.

As we can see, we have marked the moment where the camera took a picture. In that moment, the camera had the faculty to save the image on its DRAM, when it had been realized the Neural Network algorithm was applied. For that, the system has the possibility to decide when to finish the process. It is necessary to have this control because of it saves time to the operator and he can replace and verify the corresponding board.

3 Image Acquisition, processing and Segmentation

As we have been talking, the image acquisition process was realized using a VCM40 camera; the pneumatic machine counts with a little door above the place where the electronic board is inserted. This door can be closed, and then we can obtain a black area where the picture can be tacked. Figure 4 shows the structure of the pneumatic machine and the VCM40 camera position into it.

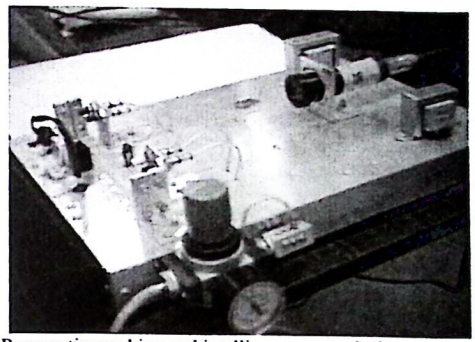


Fig. 4 Pneumatic machine and intelligent camera before to start proves.

When the electronic board that contains the displays is collocated in its position, the process begins and the following aspects must be considered:

a) The process finished satisfactorily
b) The process is stopped because of any trouble was identified in any components of the electronic board.

The second aspect mentioned above can be because:

a) The segments that are in the digits of the electronic board are light off.
b) The intensity of someone segments of the display is not the same like others.
c) The sound emitted for the clock is not operating adequately.

In the experiment we used a threshold to determine what kind of image was obtained from the camera. It was calculated of experimental form, the expression utilized is:

$$\hbar(t) = \begin{cases} 0 & if \quad \hbar(t) < \alpha \quad for \quad t > 0 \\ 1 & if \quad \hbar(t) > \alpha \quad for \quad t > 0 \end{cases} \qquad (1)$$

For edge detection we used Sobel operators. As we know, they are used to find the edge detection in the domain of two-dimensional image, using the gradient and Laplacian expressions that are defined as [5]:

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{2}$$

The magnitude of the gradient is given by:

$$\nabla f = |\nabla f| = \left[G_x^2 + G_y^2\right]^{1/2} \tag{3}$$

And the direction of the gradient vector is then:

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_x}{G_y}\right) \tag{4}$$

In order to apply these operators, the function *f(x,y)* must be a closed form expression which we can differentiate analytically. Clearly, we can be able to approximate the partial derivatives and over an array of constant data representing the intensities of an image. Referring at this trouble, the most obvious and straightforward way to accomplish this is simply as follow:

$$\begin{aligned} G_x &= Z_5 - Z_8 \\ G_y &= Z_5 - Z_6 \end{aligned} \tag{5}$$

The Sobel operators may be useful due their smoothing effect. The discrete operators corresponding to the partial derivatives are given by:

$$\begin{aligned} G_x &= Z_7 + 2Z_8 + Z_9 - (Z_1 + 2Z_2 + Z_3) \\ G_y &= Z_3 + 2Z_6 + Z_9 - (Z_1 + 2Z_4 + Z_7) \end{aligned} \tag{6}$$

And the corresponding masks as:

$$\begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \& \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \tag{7}$$

Figure 5 shows any image before and after to apply the Sobel mask to one of our image obtained. Also, we can see the forms of the image during the stages of the preprocessing and segmentation.

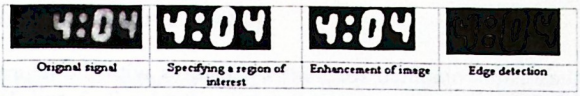


Fig. 5 The pre-processing stage on the image obtained from the intelligent camera.

4 Neural Network and Experimental Methodology

Neural Networks (NNs) are used for a wide variety of image classification tasks [6]. An object is represented by a number of features that form a d-dimensional feature vector x within an input space $X \subseteq R^d$. A classifier therefore realizes a mapping from input space X to a finite set of classes $C = \{1,.....,l\}$. A neural network is trained to perform a classification task from a set of training examples $S = \{(x^\mu, t^\mu), \mu = 1,..., M\}$ using a supervised learning algorithm. The training set S consists of M feature vectors $x^\mu \in R^d$ each labeled with a class membership $t^\mu \in C$. The network typically has as many outputs as classes and the target labels are translated into l-dimensional target vectors following a local unary representation. During the training phase the network parameters are adapted to approximate this mapping as accurately as possible (unless some techniques, such as early stopping, are applied to avoid over-fitting). In the classification phase an unlabeled feature vector $x \in R^d$ is represented to the trained network and the network outputs provide an estimation of a-posteriori class probabilities for the input x, from which a classification decision is made, usually an assignment to the class with maximum *a-posteriori* probability [7][8].

In this work, we used an auto-associative neural network based in Hopfield idea. The network consisted of a weight matrix with 4000 neurons for the image recognition because of the images obtained had 50x80 pixels. The purpose to have this neural network is to recognize the eight important elements (numbers: ***0, 1, 2, 5,*** and letters **E, *n, d,*** and the ***points***), showed in the different stages during prove of clocks. The task is then, to recognize in the same picture after to taken it, the digits and the points illuminated in the electronic board all right with the process of prove mentioned before. For that, inside of the flash RAM we have positioned the weight matrix obtained during training phase that represents the eight elements corresponding. To obtain these reference patterns we use a probabilistic method because of the position of the digits in the picture taken were different (but not so much). With 20 electronic boards we obtained the average position.

5 Results

We show the results obtained for a set of features with the neural network mentioned above and the averaging method purpose (tables 2, 3 and 4). We prove the experiments using 500 electronic boards and its elements inside. Each board has four digits and three points to recognize, the task was that during a complete process of verification and approbation of each them, 5 pictures were taken. Then, we have a total of 3000 elements analyzed:

Stage/pattern	0	1	2	5	E	n	d	Points
1	2	1	1					2
2		2		2				2
3	2	1	1					2
4	1	1						3
5					1	1	1	
Total	5	5	2	2	1	1	1	9

Table 2 Number elements analyzed and its corresponding stages.

Then, with 500 electronic boards we have the following results for each pattern:

Stage/pattern	0	1	2	5	E	n	d	Points	
1	1000	500	500					1000	
2		1000		1000				1000	
3	1000	500	500					1000	
4	500	500						1500	
5					500	500	500		
Total	2500	2500	1000	1000	500	500	500	4500	13000

Table 3 Total of elements analyzed.

The results obtained for each pattern according the stage were the following:

Stage/pattern	0	1	2	5	E	n	d	Points	
1	980	491	496					994	
2		983		991				994	
3	980	491	496					994	
4	488	495						1438	
5					494	490	486		
Total	2448	2460	992	991	494	490	486	4450	12881

Table 4 Results obtained after to apply the Neural Network to the industrial process.

Then, with the results obtained we reached a recognition rate of:

$$\%recognition = \frac{12781}{13000} x100 = 98.5461$$

The high performance of the neural network demonstrated in the percent recognition avoid to have so much electronic boards in the market, because of the number of rejected pieces in this experiment were 121 elements (the worst in the analysis effected), another experiments were realized with other set of 500 electronic boards and the best only rejected 25 electronic boards. The troubles were because the dip switch was damaged or the electronic board was not welded adequately.

6 Conclusions and future works

In the present paper we demonstrate an application of the neural networks in an industrial process that counts with a set of stages predetermined. Also, we introduced a method to control the limitations that we had at using this type of classifiers in these applications. The primordial aspects used were:

- Image acquisition. Using the VCM40 (an intelligent camera manufactured by Vision Components),
- Pre-processing of image. Applying Sobel masks to edge detection,
- Segmentation. Using an empirical threshold; obtained from the pictures.
- Training and recognition. Using a neural network; for that.

The value of the threshold was determined of empirical form, it was obtained after that we analyzed 20 electronic boards that was used also for the training phase. The results that we obtained showed us that the application of the neural networks in an industrial process with more stages is possible. So verification and tested of digital clocks too. The results obtained optimize and automate a process that was realized using other technique (manually); results demonstrate that.

In order to increase the classification performances in this application, there are several actions that can be attempted. First, we must try realizing the recognition phase using another neural network and comparing the results obtained with that we reported here. So, this is a project in process; because the results reported are the first experiments after that the pneumatic machine and the digital clocks were assembled. For that, only we had presented these results. Second, a new neural network structure can be proposed to avoid using to set of electronic boards choosing the average position of them. Neural networks like LAM, BAM, ALFA, BETA, etc., can be used.

Finally, the algorithm and the approach implemented here will be applied to another kinds of digital clocks manufactured by ELTEC, where the complexity is different because of not only neural network must classify the set of digits, but that neural network must recognize what digital clock type is presents and the subcategory that it pertains.

References

[1] Pope A. R. ***Model-Based Object Recognition. A survey of recent research***. University of British Columbia, Vancouver, Canada, Technical Report 94-04, January 1994.

[2] Singh S., Markou M., Haddon J., ***Detection of new image objects in video sequences using neural networks***. Proc. SPIE Vol. 3962, p. 204-213, Applications of Artificial Neural Networks in Image Processing V, Nasser M. Nasrabadi; Aggelos K. Katssagelos; Eds., 2000.

[3] Fay R., Kaufmann U., Schwenker F., Palm G., ***Learning objects recognition in a neurobotic system.*** In: Horst-Michael Grob, Klaus Debes, Hans-Joachim Böhme (Eds.) 3rd Workshop on Self Organization of AdaptiVE Behavior (SOAVE 2004). Fortschritt-Berichte VDI, Reihe 10 Informatik / Kummunikation, Nr. 743, pp. 198-209, VDI Verlag, Düsseldorf, 2004.

[4] Wang W., Zhang A. and Song Y. ***Identification of objects from image regions.*** IEEE International Conference on Multimedia and Exp (ICME 2003), Baltimore, July 6-9, 2003.

[5] Abhijit S. Pandya, Robert B. Macy. ***Pattern Recognition with Neural Networks in C++.*** CRC PRESS & IEEE PRESS. A CRC Cook Published in Cooperation with IEEE PRESS, 1996.

[6] Felzenswalb P. and Huttenlocher D. ***Efficiently computing a good segmentation.*** In IEEE Conference on Multimedia and Expo (ICME 2003), Baltimore, July 6-9, 2003.

[7] Bishop C. M. ***Neural networks for Pattern Recognition.*** Oxford University Press, 1995.

[8] Nicolas Amezquita Gomez and Rene Alquezar. ***Object Recognition in Indoor Sequences by Classifying Image Segmetnation Regions Using Neural Networks.*** 10th Iberoamerican Congreso on Pattern Reognition, CIARP 2005, Havana, Cuba, November 2005, Proceedings.

Low Cost Multi-Channel Data Acquisition System with Graphical HMI

Antonio Hernández Zavala, Oscar Camacho Nieto,

Cornelio Yañez Márquez, Osvaldo Espinosa Sosa

Centro de Investigaciones en Computación – Instituto Politécnico Nacional MEXICO Av. Juan de Dios Batiz s/n UP Zacatenco, México DF.
antonioh_z@prodigy.net.mx
oscarc@cic.ipn.mx

Abstract. With the objective of reduce the breach of technology that first world countries have to manage industry's process control, in this paper a low cost multi-channel data acquisition system integrated with a human-machine interface HMI is developed. To validate the efficiency of our MCDAS system with regard to conventional systems, as case of study a tank of water with temperature sensors and a heater are used to supervise and modify its state with multiple parameters to read from an acquisition card developed. Finally it was detected that our hardware system, that has integrated a 10 bit, 16 channel analog to digital converter, a Harvard architecture RISC processor at 20 MHz and RS-232 serial transmission, is a proprietary technological development which is versatile and is realized at low cost, and will be used for more cases of study at the institute.

Keywords. SCADA, Data Acquisition Card, Control, Supervision.

1 Introduction

In real-time processes at industry, low cost supervision systems and reliable data management on an easy of use interface are required. Almost all industries have a computer on their hands, and their users are familiarized with graphical environment because its simplicity to visualize changes on the environment of the questioned variable, and respond to the changes given by the user. A system that allows a direct interaction with real processes by means of a desktop computer is commonly called SCADA (System of Control And Data Acquisition).

SCADA's are not new at industry field, there is too much work done, since small systems dedicated to a single process up to big ones that allow the monitoring of a complete production plant. Among SCADA developers, we have National Instruments with LabView software, which have all the necessary tools for signal conditioning, data acquisition modules and its graphical friendship programming language able to

© L Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 85-93

interconnect its wide catalog of products. This is a very complete product line which can be adapted to any real process, but software and hardware modules are very expensive and ties users to their specific technologies as happen with the rest of commercial developers.

A good alternative to have an SCADA for monitoring a process, at low price, is to create the complete platform for that specific process; letting options for more cases, by using a robust general purpose object oriented programming language for the creation of an HMI for the process viewing and data transmission. Semiconductors are also required to obtain a binary value that represents that real world variable. Their low price is added to the advantages that represent the creation of this system. The acquisition card uses a general purpose microcontroller, ADC, DAC and signal conditioning that uses solid state circuits

The system and the process of schema on Fig. 1, are a tank of water with six temperature sensors placed on different points, they are signal conditioned prior to analog to digital conversion, and read by a general purpose microcontroller [2] which performs the task of send the sensors measures to the PC, via serial communication. The PC contains the HMI that is dedicated to plot measures and modify the environment of the variable. On Fig. 2 is a picture of the physical system.

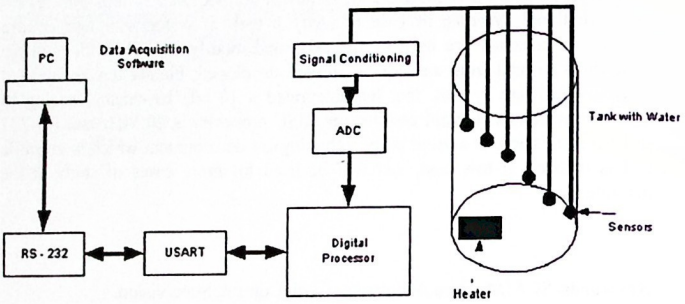


Fig. 1. Block diagram of the system.

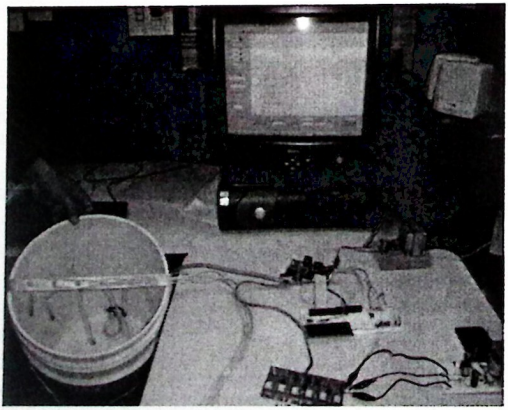


Fig. 2. Physical Prototype System.

This paper gives a description about the realization of all the modules of a SCADA system which is intended to be used as measurement and controlling system for different experiments at research centre. The rest of the document is divided in four parts: following part shows how translation of water temperature into a comprehensive voltage or signal conditioning stage, this includes the hardware used for sensing and the 10 bit analog to digital converter. On the next part digital data treatment is discussed. Fourth part is dedicated to the HMI characteristics on a general purpose programming language, in this case C++Builder^{R}. Finally at fifth part results and conclusions are reviewed.

2 Signal Conditioning

The physical variable to be represented is temperature, in order to obtain a voltage level representation at low cost, six LM335 temperature sensors with calibrated output according to manufacturer data [1][5] were used; the resolution of the signal given is 10mV per grade in a linear behavior; calibration is at 2.98V when temperature is 25°C, this values are used as reference ($Vout_{T0}$, T_0) for:

$$Vout_T = Vout_{T_0} \times \frac{T}{T_0} \tag{1}$$

Here $Vout_T$ is the voltage level obtained for a temperature T using the previous reference values. The output of the sensor is connected to an operational amplifier voltage follower [4], which gives us an output for each channel as in Fig. 3.

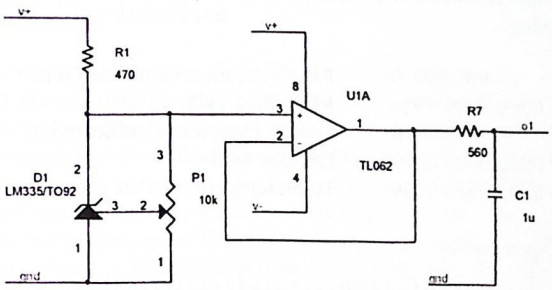


Fig. 3. Signal conditioning for a single temperature sensor.

In order to realize the digitalization of the variable, a 10 bit successive approximations ADC embedded in the COP8 MCU is used. It takes 15 clock cycles to have a sample and the result is available at the ADRSTH and ADRSTL 8-bit registers [2]. Since a 20MHz clock frequency for the microcontroller is used, it takes 0.75us to get one sample. This can be configured to different sampling rates by its internal registers. Depending on the number of channels that are read, sampling time gets slower while number of channels increase.

3 Digital Decoding of Data Packages into Comprehensive Data

This part is the most important because if data is not correctly decoded, there will be great mistakes on readings as on Fig. 4 which shows (a) bad ordered data for ambient temperature and (b) noise from electricity on water. To correct (a), the following algorithm is proposed, for (b), a low pass antialiasing filter is used.

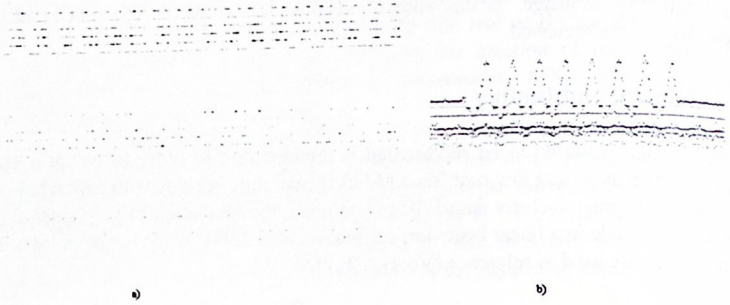


Fig. 4. Disorders at data decoding, a)Disordered data decoding, b)Noise affecting measures

A sample of each channel is 10 bit length so it does not fix on a byte, then two bytes are used as they are read from converter registers ADRSTH and ADRSTL, the first contain the 8 most significant bits, the second contain the 2 least significant bits in the upper 2 bits [6]. Then, a sample is on a single word size, for six samples per channel there are 12 bytes per sample pack and when received, the decoding process is like the following:

9999 9999	First byte, most significant bits.
0000 0099 9999 9900	Multiplied by 4.
9900 0000	Second byte, less significant bits.
0000 0000 0000 0099	Divided by 64.
0000 0099 9999 9999	10 bit temperature data on a single word.

That can be expressed as:

$$swDATA = (HIREG * 4) + (LOREG / 64) \tag{2}$$

This is the task that will be applied to every word in the package received and will loop until no data received. On Fig. 5a and 5b the flow diagram of the timer interrupt is shown divided in two parts: the first part, is the beginning, where some possible fails on data reception are handled in a way that let us work; on part *b* a loop until no data available is executed for data decoding with (2); plotting of the points and unit visualization are realized with (3), where temperature from a single word data is obtained at °C and visualized on the graphics screen of HMI.

$$T°C = \frac{swDATA * 298}{800} - 273 \quad (3)$$

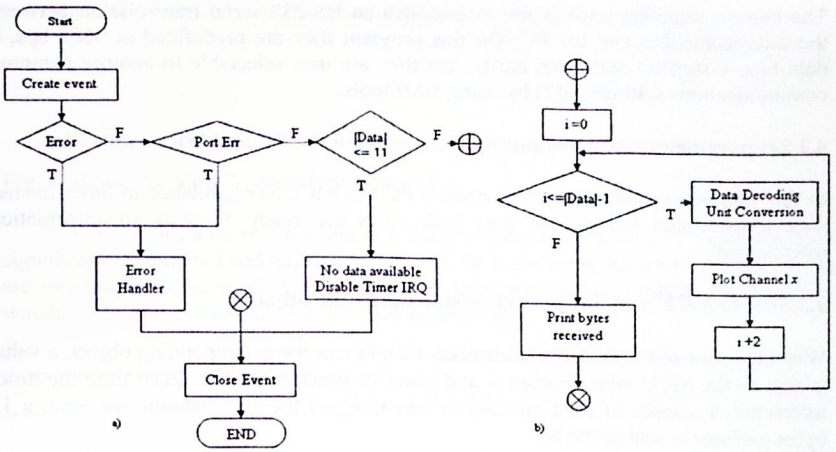


Fig. 5. a)Failure detection, b)Data decoding until no data available

4. Human Machine Interface

HMI Software was developed on C++ Builder 5.0™, which is a robust graphical environment for general purpose Object Oriented Programming. In this language, when something is done to an object at runtime, an event is generated and calls its associated subroutine, which can in turn do the desired task [8][9].

The program works something like this: at start up it initializes all parameters required and wait until something happens. First step is for the user, who must ensure that the MCU is reset. Next, to set the communications parameters double click on the respective object. Then a transmission can be started, and the plot of the curves representing the actual state of the variables is seen on screen graphics. Transmission will loop until the user stops it.

While a transmission is running, the user can save on a file, previously created with the HMI, the measures obtained on a given time period. The software is able to do the following tasks that are detailed later:

1. Set parameters for communications between the PC and microcontroller.
2. Send to MCU a signal to start or stop data transmission.
3. Decoding of data packages into comprehensive data.
4. Changes number of samples per second, from 3 to 50 per channel.

5. Saves on file a desired segment of signals.
6. Reproduces signals previously recorded for analysis

4.1 Communications Settings

The communications settings are to establish an RS-232 serial transmission between the microcontroller and the PC. On this program they are predefined at 9600 bps, 8 data bits, 1 stop bit and Even parity, but they are user selectable to another common communications settings[2][7] by using HMI tools.

4.2 Set parameters for Communications between the PC and Microcontroller

In this function a connection line between PC and MCU is established in order to start data transmission being sure that both sides are ready to start an information interchange.

4.3 Sent to MCU a signal to start or stop data transmission

When the user wants to start a transmission and press the corresponding object, a value is sent to the MCU who decodes it and starts its timer execution. Each time the timer interrupts, a sample of each channel is taken. When the six channels are read, a 12 bytes package is sent to the PC.

On the PC, a timer which generates an interrupt every time period given is activated and makes transmission a loop of port buffer readings until the user stops it or no more data available.

When the user stops transmission another byte is sent to MCU and indicates that the timer must be stopped and values reset.

4.5 Saves on file a desired segment of signals

The user can choose to record some signals on a time interval by selecting *Record signal* on the toolbar; a file must be first created with HMI [8][9]. This procedure can be stopped when the user wants it.

4.6 Reproduces signals previously recorded for analysis

A signal which has been recorded can be reproduced and viewed later, with no connection to the acquisition card, or some signal processing algorithm can be applied to test signal state without affecting real system [8][9].

5. Results and Tests

Several tests have been realized since the beginning of the project, we encountered many fails and correct them. All measure plots presented here were performed by this HMI software, note that some plots are reproduced from file and its measures have an offset, which is intended for analysis and visualization but it does not seen on real time plots, if they have the same measure, they are overlapped.

On Fig. 6 there is a failure, this occurs when the heater is turned on, and it reduces voltage on the whole circuit. Therefore it was supplied independent from the digital circuitry.



Fig. 6. Heater Stealing Voltage (reproduced).

Fig. 7a shows a correct reading at a uniform temperature of 22°C, note that these signals are reproduced and have an offset. Fig. 7b is the same but on real time, all lines are overlapped because all of them have the same temperature, except for lower line which is out of the water and has ambient temperature.



a) b)

Fig. 7. Ambient temperature measures. a) Reproduced, b) Real time.

When the heater is on, water temperature is increasing its temperature from the effect of the heater on it, which can be seen on Fig. 8, note that the higher temperature correspond to the sensor which is closer to the heater, the lower line comes from the sensor which is more distant from heater.

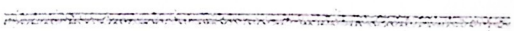


Fig. 8. Measures when heater is on.

6. Conclusions

With the development and construction of the MCDAS system (Fig. 1) with HMI, proprietary technology was developed at low cost. About the programming language, this option is still expensive, but it is very versatile and every program made can be scaled up by using all the tools that language can provide, besides this, the

communications protocol can be proposed as simpler as user wants or application requires. By the hardware side, great cost savings are reached, this is because the whole platform was realized since the beginning using low cost parts, but we are still looking for a cheaper and robust microcontroller and we have in mind that for great volumes, cost gets reduced.

With the correct configuration of used semiconductors, as a case of study we used a tank with water on which temperature was monitored using sensors on different places at real time. This is able of change water temperature by a heater. We found that the system effectively keep track of the variable behavior and can affect the state of the system easily.

With this project, we have a monitoring tool that can be used with other industrial processes previously signal conditioned to operate from 0 to 5 Vcc, reducing considerably the hardware acquisition cost. In fact, this tool will help us to have industrial equipment as on first world countries but at lower price and totally designed by us.

This project is still on growing development, we are looking for the addition of new tools that allow an easier use and understanding of the referenced process.

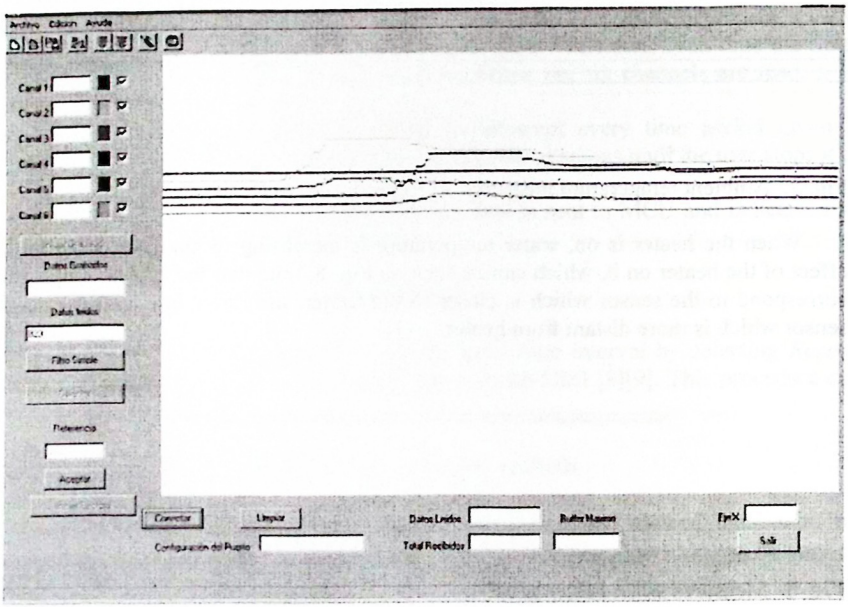


Fig. 9. Graphical interface for MCDAS.

References

1. LM135/LM235/LM335, LM135A/LM235A/LM335A Precision Temperature Sensors. National Semiconductor Datasheet. DS005698 (2000)

2. COP8CBR9/COP8CCR9/COP8CDR9 8-Bit CMOS Flash Microcontroller with 32k Memory, Virtual EEPROM, 10-Bit A/D and Brownout. User Manual National Semiconductor. DS101374 (2002)
3. Hernández A., Algoritmo de Lógica Difusa para Corrección de Error en Aplicaciones de Tiempo Real, MC Tesis, Centro de Investigaciones en Computación IPN, México, 2004.
4. Boylestad, N.: Electrónica: Teoría de Circuitos y Dispositivos electrónicos. Prentice-Hall. 8th. Edition. (2003)
5. Maloney, T. J.: Electrónica Industrial Moderna. Prentice-Hall. 5ª Edicion. (2006)
6. Tocci, R. J.: Sistemas Digitales, principios y aplicaciones. Prentice-Hall. 6th Edition. (1996)
7. Denver, A.: Serial Communications in Win32. Microsoft Windows Developer Support. (1995)
8. Hollingworth, J., Butterfield, D., Swart, B., Alsop, J.: Borland C++ Builder 5.0 Developers guide. SAMS Publishing. (2001)
9. Schildt, H.: Borland C++ Builder: The Complete Reference. Mc Graw Hill. (2001)

Patterns Recognition for Shrimp Aquaculture

Luis P. Sánchez, José Juan Carbajal Hernández and Oleksiy Pogrebnyak
Center for Computing Research, National Polytechnic Institute
Av. Juan de Dios Batiz s/n casi esq. Miguel Othon de Mendizabal, Col. Nueva Industrial Vallejo. CP 07738. Mexico City, Mexico
lsanchez@cic.ipn.mx

Abstract. Mexico is one of the main exporters of shrimp at international level. The environmental quality in prawn ponds is essential for good production of species. If there exists any change in one of the environmental parameters, and this is not early controlled, it could generate instability into the pond, the organisms will be stressed and as a result they can get sick, in some cases, the sickness could be mortal. Nowadays, there is not software that attacks this point. In this work, we create environmental variables patterns, such as temperature, salinity, pH, dissolved oxygen and turbidity, and they are recognized by a feedforward neural network. This patterns recognition makes possible to diagnose and to present the trend of shrimp artificial habitat behavior and to take actions of control and prevention.

1 Introduction

In the coastal zones, the fishing contributes to generate important gains. The production can be affected by environmental factors, such as tropical storms, cyclones or diseases. Artificial sowing of the species tries to intensify the amount of the product and the quality in the same one.

Mexico is considered as the fourth producer at international level. Shrimp artificial sowing is not simple, this species is very sensible and any change in its habitat as much produces great negative effects in its growth as in its reproduction. The farms that are dedicated to the culture of the shrimp have exclusively dedicated laboratories to control the environment to generate a production greater.

The shrimp aquaculture in Mexico and the world has not advanced computational systems to supervise the artificial habitat of the farms and laboratories. A computational system of this type helps significantly to improve the environmental conditions and to elevate the production and its quality [1], [2].

The main idea of this work is the creation of a system, which can help the biologists to recognize patterns of problems and their evolution in shrimp aquaculture, and thus to respond with greater rapidity against the negative effects.

Bad control on the shrimp artificial habitat produces organisms with the high stress and as consequence losses in their defenses. It generate low nutrition, low reproduction or worse still, they prearrange to acquire lethal diseases. The proposed system helps to control this problem [3].

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 95-104

The established hypotheses are the following ones:

1. It is possible and necessary to establish patterns of environmental variables with nonlinear relationships to classify the artificial habitat state of shrimp.
2. A feedforward neural network, with a linear activation function for output neurons, allows to recognize and to quantify the shrimp artificial habitat state, which is given by environmental variables patterns.

2 Technical Support and Schemes of Operation

The main environmental variables are measured of automated way or introduced to a computer by the operator, and these are the following ones:

1. Temperature (°c)
2. Dissolved oxygen (mg/L)
3. Salinity (g/L or ppt)
4. pH
5. Alkalinity (mg/L $CaCO_3$)
6. Turbidity (cm)
7. Total ammonium (mg/L)

A distributed system [4], [5] is used in the automated measurement (pH, temperature, Dissolved oxygen, etc.). An operator makes laboratory measurement (Salinity, Alkalinity, Total ammonium, etc.). The Fig. 1 shows an installation for shrimp aquaculture and the Fig. 2 presents a simplified scheme of the distributed system.

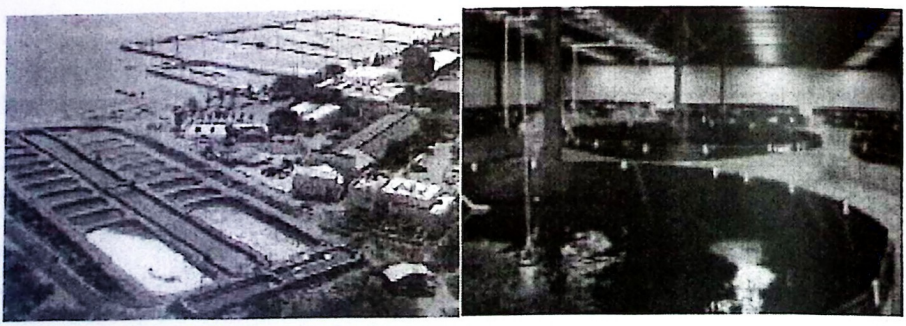


Fig. 1 Example of an installation for shrimp aquaculture

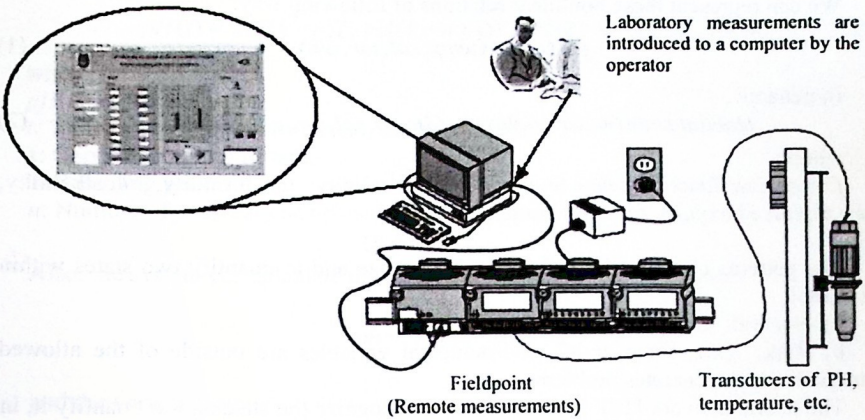


Fig. 2 Scheme of the distributed system

3 Patterns of environmental variables with nonlinear relationships

Environmental variables as pH, temperature, salinity, dissolved oxygen and turbidity have an important effect in the suitable growth of the shrimps and influence in their health. For example for Shrimp Maturation and Spawning [6], [7] almost all hatcheries require availability of oceanic-quality water on a 24-h basis. Salinity and temperature are the most important water parameters impacting production of shrimp in the hatchery, and must be maintained in a narrow range, between 27 and 36 ppt salinity and 28 C (82 F) plus or minus two degrees for most penaeids. These and other important factors in the maturation and spawning of penaeid shrimp are discussed in detail by Treece and Fox [8].

Parameters for tropical shrimp maturation and allowable ranges/24 hr [6]

Salinity:
27-36 ppt +/- 0.5
Temperature
28 C +/- 2 (80.5-84.2 F)
pH
7.8 +/- 0.2
Light
14 L, 10 D
D.O. (Dissolved oxygen)
5 ppm

These values allow to create normal patterns. In general, these variables have nonlinear relations, which have been observed and proven experimentally, although

the equations that represent them have been formulated, which is very difficult [9].

We can represent these nonlinear relations of following way:

$$pH(t) = f(temp, sal, tur, alk) \quad (1)$$

In general:

$$Habitat\ state\ (water\ quality) = f(temp, sal, tur, alk, do, ta, ...) \quad (2)$$

where:

f: nonlinear function; *temp*: temperature; *sal*: salinity; *tur*: turbidity; *alk*: alkalinity; *do:* dissolved oxygen; *ta*: Total ammonium.

The patterns classification is made to recognize and to quantify two states within the pool:

a) **Normal**: Everything is well.

b) **Risk**: One, some or all environmental variables are outside of the allowed interval, which generates problems.

The neural network [10], [11] will have to recognize the state and to quantify it, in others words, how normal or risky it is, which allows to find trend of the water quality. The total ammonium only, can cause a danger state; therefore, this situation is found by means of conditions if-else.

In Table 1 are the environmental variables of greater importance, the values and the combinations of these, for which the pool is in a normal state, of risk or danger.

Table 1. States classification by means of the combination of environmental variables.

NORMAL	
Environmental variables	**Range**
Temperature	23 – 30 °C
Salinity	15 – 25 (mg/l)
pH	7.6 – 8.6
Dissolved oxygen	6 – 10 (mg/l)
Total ammonium	Less than 0.1 (mg/l)
Turbidity	35 – 45 (mg/l)
RISK	
Temperature	Greater than 30°
Salinity	Greater than 25 (mg/l)
pH	Less than 7.6 or greater than 8.6
Dissolved oxygen	Less than 6 or greater than 10 (mg/l)
Turbidity	Less than 35 or greater than 45 (mg/l)
DANGER	
Total ammonium	Greater or equal than 0.1 (mg/l)
Dissolved oxygen	Less than 3 (mg/l)

The method determines if a danger state exists, and in such case it gives to the present values of ammonium and oxygen. If a danger state does not exist, then the neural network recognizes and quantifies the normal state or the risk state. In the three situations, the user could consult the historical values of the environmental variables.

These historical values (for example for pH) are defined as:

$$pH(t) = pH(k-n)T, pH(k-n+1)T, pH(k-n+2)T, ..., pH(k) \quad (3)$$

where:
pH(t): Historical value.
t: Time interval for the historical, defined for user.
k: Current sampling instant.
T: Sampling period.
n: Number of necessary samples for the historical, so that:

$$(n+1)T = t \quad (4)$$

Also, the system can calculate the historical average and variance, as:

$$\overline{ph} = \sum_{i=k}^{k-n} ph(i)/(n+1) \quad (5)$$

and the variance, as:

$$Vph = \sum_{i=k}^{k-n} (ph(i) - \overline{ph})^2 /(n+1) \quad (6)$$

The user can solicit a trend analysis, linear or polynomial [12].

The classification of the real measurements in training patterns was made in farms of test in several periods of sowing, and 400 patterns were used, which represent different levels of normal state and different levels of risky state.

Total ammonium is considered more danger than others variables. A minimum concentration of ammonia is enough for creating a danger state. The system recognize a danger state with this variable only when the maximum level is reached, this is done using an if-else programming and not with a neural network.

4. Neural network

Fig 3 presents the topology of better results for this application. The neural network also is able to give a deviation of the centre of the state, in other words, the state what so normal or risky is. This deviation allows to have historical of the classifications, same that make possible to establish a trend of the evolution of the water quality or artificial habitat, and the technologists can control better at shrimp habitat.

The neuron that recognizes the normal state gives the value of 1 if the state within the pool is best, if this state is deteriorated, then neuron will give inferior values to 1, and will go in reduction. In the same way, the neuron that recognizes the risk state gives the value of -1 when the neuron of normal state is in 1, and will go in ascent if the deterioration state increases. If the state within the pool is very deteriorated, then the classification will be of risk and the neuron that recognizes the risk will give a value of 1 (see Fig. 4). The Fig. 5 presents the performance in neural network training.

The Table 2 has the test results of patterns recognition for values combinations of environmental variables. This test was successful.

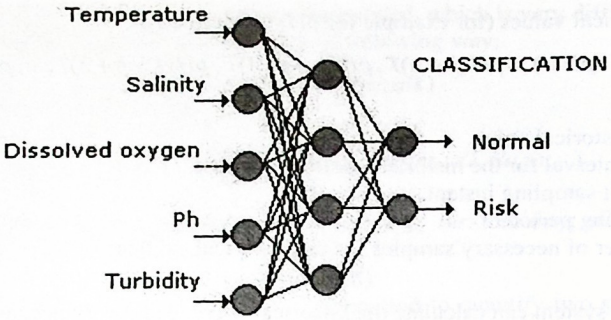


Fig. 3 Feedforward neural network

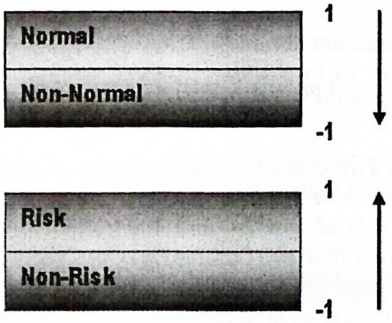


Fig. 4 Operation of output neurons with linear activation function

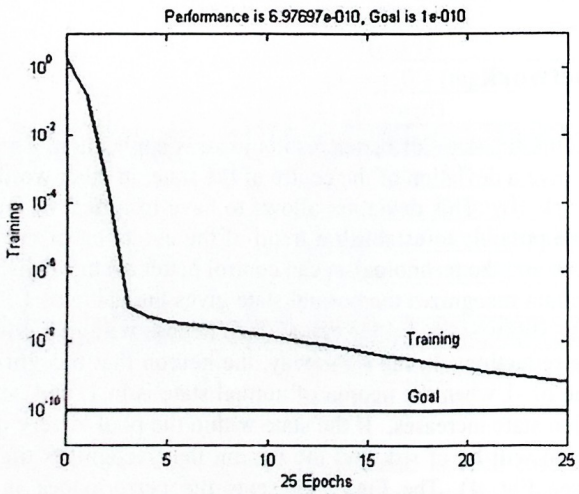


Fig. 5 Performance in neural network training

Table 2 Some tests of patterns recognition made by neural network.

Tempe-rature(^{0}C)	Salinity (mg/l)	Dissolved oxygen (mg/l)	pH	Turbidity (mg/l)	Normal	Risk
33	15	15	9	35	-1	1
23	25	6	8.2	40	-1	1
32	20	5.3	7.5	38	-1	1
40	32	5.5	9.2	39	-1	1
30	25	9	7.8	41	-1	1
27	20	8	8.5	40	-1	1
28	31	4	7.8	17	-1	1
28	28	3	7.9	12	-1	1
28	23	3	8.9	15	-1	1
27	26	2	8.1	15	-1	1
28	31	3.2	7	11	-1	1
28	34	2.3	6.5	17	-1	1
28	32	2.2	8.5	18	-1	1
23	35	3	8	11	-1	1
33	34	9	9	48	-0.60357	0.87094
37	32	15	8.3	48	1	-0.99741
34	33	7.3	8	37	-1	1
37	35	7.2	8.7	24	-1	1
35	35	11	5	24	-1	1
35	33	15	8.5	18	-1	0.81304
35	31	15	7.8	17	-1	0.87243
34	28	15	7.9	12	-1	1
34	23	15	8.9	15	-1	1
35	26	15	8.1	15	-1	1
32	31	15	7	11	-1	1
33	34	15	6.5	17	-1	0.83643
33	32	15	8.5	18	-1	0.80734
30	35	15	8	11	-1	1
33	15	8	9	35	-1	0.86891
23	25	6	8.2	40	-1	0.76977
32	20	5.3	7.5	38	0.68359	-1
40	32	5.5	9.2	39	-1	1
30	25	9	7.8	41	1	-0.99741
27	20	8	8.5	40	1	-0.99741

5. Analysis of results

The patterns classification given by the environmental variables confirms its nonlinear interrelation. The pool state cannot be determined only by a environmental variable, this is determined due to the variations of a set of them, therefore, a trained feedforward neural network represent suitably the nonlinear relations between the environmental variables.

Figures 6, 7, 8 and 9 were obtained using as reference the Table 2, and present four examples of the variations of environmental variable of individual way in axis X,

versus axis Y that presents the recognized state, by the neural network, using the values of the set (pattern) of all environmental variables.

No any environmental variable by itself defines the state of the habitat, for example, in Fig. 6, for a normal salinity of 24.5 mg/l, the neural network recognizes a normal state; for a normal salinity of 23,5 mg/l the neural network recognizes a moderate risky state.

The presented examples allow to understand the nonlinear relations between the environmental variables. These nonlinear relations are represented suitably by feedforward neural network.

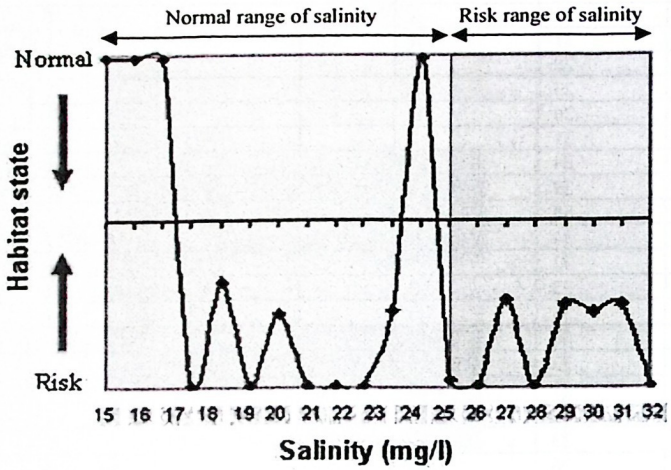


Fig. 6 Salinity variations and recognized habitat state

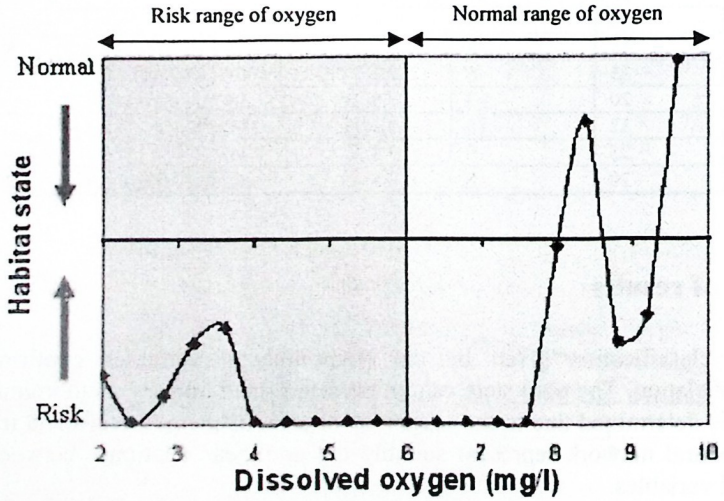


Fig. 7 Dissolved oxygen variations and recognized habitat state

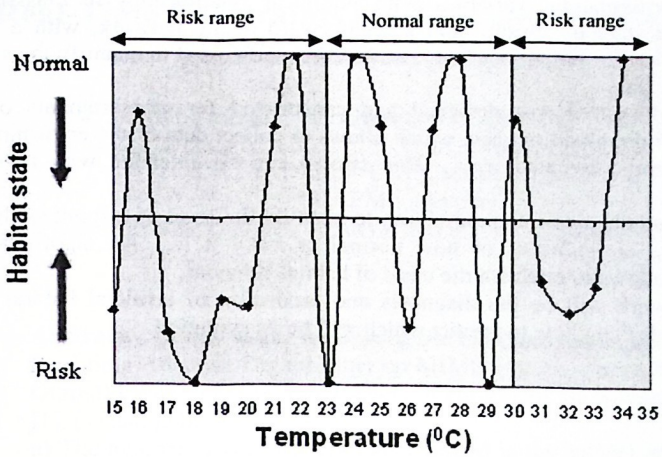


Fig. 8 Temperature variations and recognized habitat state

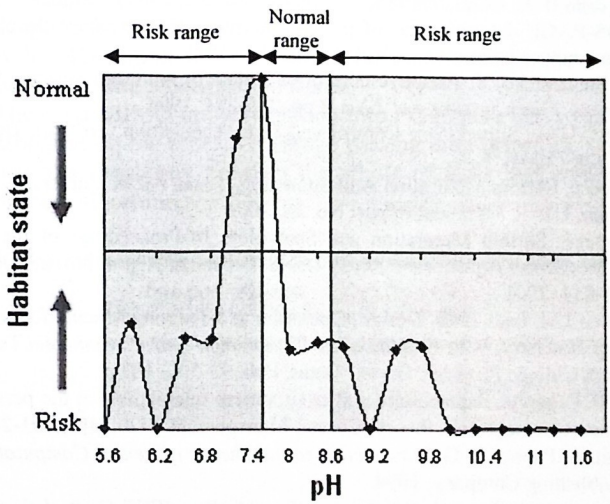


Fig. 9 pH variations and recognized habitat state (normal: 7.6-8.6)

Conclusions and Future Work

In the shrimp aquaculture, it is difficult to maintain a similar water quality to natural environmental conditions. This work allows to diagnose any problem early, being able to make an opportune treatment and to obtain harvests of better quality.

The hypotheses were demonstrated: It is possible and necessary to establish

patterns of environmental variables with nonlinear relationships to classify the artificial habitat state of shrimp, and a feedforward neural network, with a linear activation function for output neurons, allows to recognize and to quantify the shrimp artificial habitat state.

A distributed system was designed and constructed for measurements of pH, temperature and dissolved oxygen, which allows to collect data of the environmental variables, of semi-automated way. This system can be extended with relatively simple changes.

The developed diagnosis method allows to quantify the level of recognized habitat state, presenting an indicator of how normal or risky it is. An algorithm was developed that allows to establish the trend of habitat behavior.

The future work will be the diagnosis and prediction of artificial habitat state, adapting methods that allow to predict which will be its evolution.

References

1. Martínez, L., R., "Ecología de los Sistemas Acuícolas", Ed. ADT S.A., 1998
2. Arredondo, J., y Ponce, J., "Calidad Del Agua en Acuicultura", Ed. AGT S.A., 1998
3. Díaz, M., Moyano, F.J., García-Carreño, F.L., Alarcón, F.J. and Sarasquete, M.C., 1997. Substrate- SDS-PAGE determination of protease activity through larval development in sea bream, Aquaculture international, Vol. 5, Pags 461-471.
4. Su C. Y. and Stepanenko, Y. Adaptive control of a class of nonlinear systems with fuzzy logic. *IEEE Trans. Fuzzy Systems Vol 2 No. 4* pp. 285-294. 1994
5. Boyer, S.A.: SCADA: Supervisory Control and Data Acquisition. 3rd edn. Book News, Inc., Portland, OR (2004)
6. Granvil D. Treece, Shrimp Maturation And Spawning, Texas A&M University, Sea Grant College Program. UJNR Technical Report No. 28, 2001
7. Granvil D. Treece, Shrimp Maturation and Spawning. In Proceedings of the 28th US-Japan Natural Resources Aquaculture Panel. UJNR Technical Report No. 28. pp 121-134. TAMU-SG-01-814. 2001
8. Treece, G.D. and J.M. Fox. 1993. Design, Operation and Training Manual for an Intensive Culture Shrimp Hatchery, with Emphasis on *P. monodon and P. vannamei.*Texas A&M Univ., Sea Grant College Program, Bryan, Texas, Pub. 93-505. 187 p.
9. D. Bailly and P. Paquotte. Aquaculture and environment interactions in the perspective of renewable resource management theory. Coastal Management 1996;24(3):251-269
10. Hertz, J. Krogh. A. Palmer R. G. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, 1994.
11. Psaltis, D. et al. A multi layered neural network controller. *IEEE Control System Magazine, 8(3)*:17-21. Abril 1988.
12. Oppenheim, A.V., Schafer, R.W. and Buck, J.R., *Discrete-Time Signal Processing*, 2nd Edition. Prentice-Hall Int. Ed. 1999.
13. Juricek, B., Dale E., Seborg, D., Larimore, W.: Predictive monitoring for abnormal situation management. Journal of Process Control, Vol. 11 (2001) 111-128
14. Södertröm, T., Stoica, P.: System identification. Englewood Cliffs, NJ, Prentice-Hall (1989)

Real-Time Constraints for Adaptive Digital Filters on MIMO Stochastic Systems

[1,3]J. Medel-Juárez, [1,2]P. Guevara López and [3,4]A. López Chau

(CICATA[1]-DEMS[2]-CIC[3]), Mexico Z.P. 07738,
UAEM[4], Mexico
+jjmedelj@yahoo.com.mx, *pguevara@ipn.mx, *internalimix@gmail.com

Abstract. In this work, we expose the basic properties about the Real-time Adaptive Digital Filter on MIMO (Multi – Input, Multi - Output) case. This filter has a whole of conditions (in agreement to [1], [7] and [3]):
a) The number of vector operations required in the MIMO adaptive filter process by each iteration, is bounded with time constraints respect to dynamical process conditions [8] and,
b) The filter tasks vector are scheduled, and then the MIMO adaptive filter either implanted.
A MIMO digital filter is compound by two steps: the estimation parameter and identification states.
The filter MIMO implementation into PC require sets of dynamical arrays using the process message passing with priorities, without to lose the synchrony between the whole of tasks that compound its process algorithm.

Keywords: *Filter, estimation, identification, real-time, synchrony, time constraints.*

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 105-115

1 Introduction

In [1] we say that a RTDF's are used in industrial processes, control systems and monitoring systems; for example: chemical plants, manufacturing processes, airbags and fuel injection systems, voice analysis, data acquisition, medical applications, telecommunications, missile trajectories, spatial rockets, and robotics operations. The Digital Filters cannot fail, in two senses: a) response quality and b) response time, in the other sense, the processes will be crash. For these motives, the Digital Filters properties in an implementation look up as Real-time Digital Filters (RTDF): it will use into embedded systems using electronic devices or digital computers with Real-time Operating Systems (RTOS). However, all of them, require evaluating their answers with respect to desired output, these difference considered into the basic law (that bound the gains used inside of theirs) establish the high converge regions, this kind of filters is known as Real-time Adaptive Digital Filters (RTADF) in MIMO case (RTMADF).

A huge problem implementation is developed when a RTMADF is used: in practical sense, the real world is parallel (to see references in [1]) and it has many variables in interacting in it, then a Real-time Multivariable Adaptive Digital Filter practically is required, if and only if the time evolution corresponds basically to the time system evolution. In this case, generality the users observes other problem, when the computing resources are limited in operations or in time output interaction considering a lot of variables that requires computer resources with time restrictions. The concurrent task synchronized by message passing and precedence constraints is considered if we have a processor.

2 Real-Time Multivariable Adaptive Digital Filter (RTMADF)

The class of digital filters applied into PC and that interacts with dynamic processes and emits high quality output responses with time constraints and critical synchrony, will be described as RTADF (Real-time Adaptive Digital Filters). In [2], [3], [4], [5], and [6] exposing that Real-time Adaptive Digital Filters (Kalman Filter for example) can be implanted as a RTS considering the high velocity computers and facility to express the filter in recursive form. Considering in it the times measure of all tasks around the implementation of it into PC (considering the measurable: The A/D and D/A converters, processor operations, filter algorithm and precedence constraints). In [7] was exposed the definition of Real-time digital filter, considering in it the properties previously cited in this paragraph, all in SISO (Single-Input Single-Output) representation.

The basic properties of RTADF in SISO representation is the base required for to describe a Real-time adaptive digital filter in MIMO (Multi-Input Multi Output) sense. The follows definition exposes the properties of it.

Definition 1. A RTMADF is a Multivariable Adaptive Digital Filter with time constraints in the Nyquist Multivariable sense, in according to dynamical inputs and convolution feedback law filter, synchronously interacting both, accomplishing the conditions:

a. To receive and to give responses (where $\{u(k)_i \in \mathbf{U}(k)\}$, $\{e(k)_{i-j} \in E(k)\}$ and $\{y(k)_j \in \mathbf{Y}(k)\}$, with $i, j, k \in Z^+$, are the input, the feedback and, output responses, respectively.) of a real dynamical process in synchronized time, bounded them in the Nyquist Multivariable sense (to see in simple sense: [27]).
b. To express the RTMADF in recursive form, describing on line, the physical process evolution considered,
c. To give good enough responses in time with respect to dynamical process: All vector operations required into recursive and adaptation form in each iteration system evolution, are bounded at time dynamically considering the process constraints,
d. The convergence rate value is bounded into a finite interval around the real value, converging to the upper limit of the noise interacting with the system considered.
e. A RTMADF has a concurrent tasks vector, considering that the scheduling of all instances and operations around of recursive and adaptive filter requirements are bounded by unique processor operations with respect to input and output responses desired.

A RTMADF symbolically defined as ($\{u(k)_i \in \mathbf{U}(k)\}$, $\{e(k)_{i-j} \in E(k)\}$, $\{y(k)_j \in \mathbf{Y}(k)\}$) with $i, j, k \in Z^+$, bounded each set in the Nyquist multivariable sense, considering that the sampling time is obtained as a minimal distance into the control area, symbolically is expressed as:

$$t_{sample_{MIMO}} = \min\left(\left(t_{sample}(u(k)_i) \circ t_{convolution}\{e(k)_{i-j}\}\right), t_{sample}(y(k)_j)\right), \quad (1)$$

into interval k. For example, considering the second minimum sample period into the set of all sample periods of MIMO system, then the signal with largest change velocity is lost by aliasing, it cause that filter do not describe correctly the dynamical evolution of process. It's a contradiction and, consequently is required to select the shortest sample time into the control area: Basically, this area is conformed by a set of points, with relative distance between them and absolute distance with respect to the reference point, where the minimal absolute distance is contained in all control area points.

The system should have two conditions: a) $t_{sample_{MIMO}}$ selected, is lowest Euclidian distance into the temporized control area (1), and hasn't lost information by aliasing, b) it has a $t'_{sample_{MIMO}} \subset Y(k)$ lower than $t_{sample_{MIMO}}$ selected previously in the same interval time, indexed by k. In this case, the filter doesn't have defined its control area and has negative aliasing, and the signal has perturbations, losing information.

In this case each sampling time in this kind of filters has a basic measures:
$t_{sample}(u(k)_i) = \mu[t_{sample}(u(k)_i), \overline{t_{sample}(u(k)_i)})$,
$t_{convolution}(e(k)_i) = \mu[t_{convolution}(e(k)_i), \overline{t_{convolution}(e(k)_i)})$ and
$t_{sample}(y(k)_j) = \mu[t_{sample}(y(k)_j), \overline{t_{sample}(y(k)_j)})$ (to see: [9]), where
$t_{sample}(u(k)_i) \leq 0.5(F_{\max}(u(k)_i))^{-1}, t_{convolution}(e(k)_{i-j}) \leq 0.5(F_{\max}(E(k)_{i-j}))^{-1}$ and
$t_{sample}(y(k)_j) \leq 0.5(F_{\max}(Y(k)_{i-j}))^{-1}$ in agreement to [8] and [27].

In the shortest sampling time $t_{sample_{MIMO}}$ includes the minimal sample times input set $U(k)$, sample times feedback law set $E(k)$ and sample times output set $Y(k)$. But the maximal times distance into the control area previously defined, it doesn't highest that the times evolution in the Multivariable Nyquist concept, losing synchrony, because the response times are out of the intervals indexed with k, defined by $t_{sample_{MIMO}}$ evolution delaying and displacing its interactions, obtaining a queue old problem.

Definition 2. The time constraint vectors of RTMADF is compound by the follows vectors set (to see: [10]): Arrival time ($\mathbf{I}(k)$)., Computation time ($\mathbf{C}(k)$)., Minimum relative deadline ($\mathbf{D}_{\min}(k)$): The minimum absolute deadline is $\mathbf{Id}_{\min}(k) = \mathbf{I}(k) + \mathbf{D}_{\min}(k)$., Maximum relative deadline ($\mathbf{D}_{\max}(k)$): The maximum absolute deadline is $\mathbf{LD}_{\max}(k) = \mathbf{I}(k) + \mathbf{D}_{\max}(k)$., Start time ($\mathbf{s}(k)$)., Finishing time ($\mathbf{f}(k)$)., Lateness ($\mathbf{L}(k)$)., Premature time ($\mathbf{P}(k)$)., Interaction Period (T) is $t_{sample_{MIMO}}$., Convergence time (t_c), such as $d_{c_\min} \leq t_c < d$ where $d_{c_\min}$ and, d have a minimal convergence deadline and convergence deadline, respectively.

Each time constraint vector have a set of elements such that describing to the set of tasks to build a RTMADF.

Commonly, arrival time ($\mathbf{I}(k)$), start time ($\mathbf{s}(k)$) and finishing time ($\mathbf{f}(k)$) are absolutes and the others times are relatives.

The computation time ($\mathbf{C}(k)$) defined as $\mathbf{C}(k) := \mu[\mathbf{s}(k), \mathbf{f}(k))$ and considering [9], expressed as $\mathbf{C}(k) = \mathbf{f}(k) - \mathbf{s}(k)$, where both absolute times are increasing in time evolution, i.e., $f(k) \uparrow f(m)$ *and* $s(k) \uparrow s(m)$, respectively; where the index interval described by m indicate the convergence of the RTMADF with respect to Definition 1. Nevertheless, other side, computation time vector $\mathbf{C}(k)$ either defined by the measure of all segments of time disjoints describe these in symbolic form by the set $\{C(k)_r\}$ where each $C(k)_r$ is defined in the same metric as $\mu[s(k)_r, f(k)_r)$ $\forall r \in Z^+$ accomplishing that $\sum_{i=\overline{1,n}} C(k)_i < t_{sample_{MIMO}}$. Observing that $C(k)_r = \mu[s(k)_r, f(k)_r)$ $\forall r \in Z^+$, and considering that $[s(k)_r, f(k)_r)$

$\bigcap[s(k)_m, f(k)_m) = \phi \quad \forall r \neq m$ then $\mu(\bigcup_{r=\overline{1,n}}[s(k)_r, f(k)_r) = \sum_{r=\overline{1,n}}[s(k)_r, f(k)_r)$, in other words $\sum_{r=\overline{1,n}} \mu([s(k)_r, f(k)_r)) = \sum_{r=\overline{1,n}} C(k)_r$, but $C(k) = \sum_{r=\overline{1,n}} C(k)_r$ because $\mu([s(k), f(k))) = \sum_{r=\overline{1,n}} \mu([s(k)_r, f(k)_r))$ and $\mu([s(k), f(k))) < t_{sample\ MIMO}$, i.e., $\sum_{r=\overline{1,n}} C(k)_r < t_{sample\ MIMO}$.

Finally in this section, the RTMADF has a Convergence time, expressed symbolically by (t_c) and is a maximum scalar value defined as $t_c := \max\{f(k = m_{\max})_i\}$ with $\{f(k = m_{\max})_i\} \subseteq f(m)$, such that in other case don't converge yet, considering the previous concepts, exposed before definitions.

3 Multivarible Adaptive Digital Filter considering the Instrumental Variable as Estimator and Kalman Filter as Identificator

A classical problem in that used a RTMADF is a time-varying parameter estimation, consider a real dynamic process and the necessity of estimate in Real-time its parameters (the best gains required into the Kalman filter identificator). The most immediate and simple idea is to include a discounting procedure in an estimation algorithm i.e., a procedure for discarding (forgetting) old information. The most common way to do is to introduce an exponential forgetting factor (FF) into the corresponding estimation procedure in which use an instrumental variable (to see: [11]). A large group of papers deals with constant scalar FF, for example: In [12] shown that, when a FF applied to the past data, the convergence exponentially bounded, and prediction coefficients fluctuate around the least squares estimates in the steady state, in probability sense. A small forgetting factor increases the convergence rate, but results have a larger fluctuation around of real values and convergence time is increased, to see: [13]; these authors present a theoretical analysis of the performance of the standard FF in a recursive least squares algorithm, they used it combination in the tracking of time-varying linear regression models. As application of it, [14] considers it with adaptive techniques to control the space station in orbit with moving payload.

In [15] is presented a modified sequential least squares parameter estimation method and compares its performance to that of standard least square method techniques using a simulated nonlinear F-16 with multiaxial thrust-vectoring aircraft. It is shown that the suggested method offers significant improvement in performance over the conventional least squares method parameter estimation by providing a recursive estimation algorithm that penalizes noisy estimates and is less subject to ill-conditioning as its FF is reduced.

The main characteristic of these algorithms in agreement to the set of researcher cited above ([16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26], in the others) is to permit to estimate through of scalar exponential FF, the nonstationary dynamics of system.

In [27] suggested a new approach based on the use of the recursive version of the Instrumental Variable Method (IVM) with a constant Matrix FF (MFF) providing the best estimation quality than a scalar one and espoused the properties of matrix

variable FF (VFF) according to the current prediction error. The error convergence analysis was development and evaluated with respect to least square method, both with MFF, where the IVM is approximately two times with respect to LSM, evaluated these in probability sense.

The model considered (to see: [28], [29]) is expressed in symbolic form in (2)

$$d\mathbf{x}_t = [\mathbf{A}(\omega_t')\mathbf{x}_t + \mathbf{B}u_t]dt + \mathbf{D}\xi_t \tag{2}$$

Where the vector state is expressed as:

$$\mathbf{x}_t := [i_{a_t}\ i_{b_t}\ \varphi_{a_t}\ \varphi_{b_t}] \in R^4 \tag{3}$$

The state space vector containing bounded signals

$$\mathbf{U}_t := [u_{a_t}\ u_{b_t}]^T \in R^2 \tag{4}$$

ξ_τ is defined as Standard Brownian motion process. The matrices $\mathbf{A}(\omega_t')$ and $\mathbf{B}$ and the vector $\mathbf{D}$, are defined as follows:

$$\mathbf{A}(\omega_t') = \begin{bmatrix} -\gamma & 0 & \frac{K}{T_r} & p\omega_t' K \\ 0 & -\gamma & -p\omega_t' K & \frac{K}{T_r} \\ \frac{M}{T_r} & 0 & -\frac{1}{T_r} & -p\omega_t' \\ 0 & \frac{M}{T_r} & p\omega_t' & -\frac{1}{T_r} \end{bmatrix} \tag{5}$$

$$\mathbf{B} = \begin{pmatrix} \frac{1}{\sigma L_s} & 0 \\ 0 & \frac{1}{\sigma L_s} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 0.000573 \\ 0.000457 \\ 0 \\ 0 \end{pmatrix} \tag{6}$$

Here ω' is varying according to the following dynamics:

$$d\omega_t' = K_I \mathrm{H}_\tau - K_{II} \tag{7}$$

$$K_I = \frac{pM}{JL_r},\ K_{II} = \frac{T_l}{J},\ \mathrm{H} = \varphi_{a_t} i_{b_t} - \varphi_{b_t} i_{a_t} \tag{8}$$

$$\omega_0' = 0.01 s^{-1} \tag{9}$$

All parameters participating in the description of the dynamic model (2) are used in agreement to [10]. With the follows gains:

$$K = \frac{M}{\sigma L_s L_r},\quad \gamma = \frac{\mathrm{R}_s}{\sigma L_s} + \frac{R_r M^2}{\sigma L_s L_r^2}, \tag{10}$$

$$\sigma = 1 - \left(\frac{\mathrm{M}^2}{\sigma L_s L_r^2}\right) \tag{11}$$

Selecting the time sequence t_τ as:

$$t_{\tau+1} = t_\tau + \Delta_\tau,\ \Delta_\tau = 0.015\mathrm{s} \tag{12}$$

The expression (2) has in discrete time the following space state:

$$\mathbf{x}_{\tau+1} = \mathbf{A}_\tau \mathbf{x}_\tau + \Theta_\tau \zeta_\tau' \tag{14}$$

Where $\mathbf{A}_\tau = [\mathbf{I} + \mathbf{A}(\omega_\tau)\Delta_\tau]$, $\Theta_\tau = [\mathbf{B}\Delta_\tau \ \ \mathbf{D}]$, $\zeta_\tau' = [\mathbf{U}_\tau \ \ \xi_\tau']$ With:

$$\omega_{\tau+1}' = \omega_\tau' + [K_I \mathrm{H}_\tau - K_{II}]\Delta_\tau \tag{15}$$

The random variable ξ_τ' in differences has the form:

$$\xi_\tau' = \xi_{\tau+1} - \xi_\tau \tag{16}$$

The output signal described as $\mathbf{Y}_\tau$, has in discrete time the symbolic expression:

$$Y_\tau = A_\tau Y_{\tau-1} + \theta_\tau \zeta_\tau \tag{17}$$

The problem is to generate an optimal estimation with respect to $\mathbf{A}_\tau$ at real-time in agreement to both definitions, requiring a good enough answer with respect to real modeled system. The estimator was described in [27], where the recursive form of it is expressed as:

$$\hat{\mathbf{A}}_\tau = \hat{\mathbf{A}}_{\tau-1} + (\mathbf{I} - \hat{\mathbf{A}}_{\tau-1}\mathbf{z}_\tau)\vartheta_\tau^T \Gamma_\tau,$$

$$\Gamma_\tau = \mathbf{R}^{-1}\Gamma_{\tau-1} \frac{\mathbf{R}^{-1}\Gamma_{\tau-1}\mathbf{z}_\tau\vartheta_\tau^T\mathbf{R}^{-1}\Gamma_{\tau-1}}{1 + \vartheta_\tau^T\mathbf{R}^{-1}\Gamma_{\tau-1}\mathbf{z}_\tau} \tag{18}$$

The time evaluation of the corresponding performance index J_n given by

$$J_n = \frac{1}{n}\sum_{\tau=1}^{n} tr\left\{(\hat{\mathbf{A}}_\tau - \mathbf{A}_\tau)(\hat{\mathbf{A}}_\tau - \mathbf{A}_\tau)^T\right\} \tag{19}$$

Now, the Kalman filter identificator has the following structure:

$$\hat{y}_{n/Y_{n-1}} = \hat{A}_n \hat{x}_{n/Y_{n-1}} + \theta_\tau \zeta \tag{20}$$

With the innovation process described into adaptive sequence by $\alpha_n = y_n - \hat{y}_{n/Y_{n-1}}$, the matrix $G_n = \hat{A}_\tau K_{n,n-1}\Sigma_n^{-1}$, and the adaptive error defined by the second probability moment $K_{n,n-1} = E\left[\varepsilon_{n,n-1}\varepsilon_{n,n-1}^H\right]$, considering that $\varepsilon_{n,n-1} = x_n - \hat{x}_{n/Y_{n-1}}$, $Q_2 = K_{n,n-1}$, and $\Sigma_n = K_{n,n-1} + Q_2$.

The internal identificator described in adaptive sense, has the form

$$\hat{x}_{n+1/Y_n} = \hat{A}_\tau \hat{x}_{n/Y_{n-1}} + G_n \alpha_n \tag{21}$$

The real time multivariable adaptive filtering has the optimal forgetting factor matrix based in [27]:

$$R^* = \begin{bmatrix} 0.99989 & 0.0009 & 0.0009 & 0.0005 \\ 0.0009 & 0.99898 & 0.0003 & 0.0007 \\ 0.0009 & 0.0003 & 0.9999 & 0.0002 \\ 0.0005 & 0.0007 & 0.0002 & 0.89898 \end{bmatrix} \tag{22}$$

4 RTMADF Implementation (Dc - Derivate Motor Description)

The time-varying matrix with respect to ω_τ' using the algorithm expressed in (18) with (22), and the adaptive Kalman filter using the algorithm expressed from (20) to

(22) implemented in QNX RTP Ver. 6.0 using the following hardware and software elements: **Hardware:** A D.C. Motor, with follows properties: 20 V , 1 A , 18000 rpm , two poles, with permanent field., Power unit A/D, D/A, 5 V input, 20 V output, 0.05 A input, 5 A output., PC Pentium III, 400 MHz , 64 MB ., A/D D/A card Advantech PCL 818L. **Software:** QNX RTP 6.0 Real-time Operating system. MicroPhoton Development Kit., Watcom C 10.0.

The time evaluation of the time varying parameters using (18) depicted in Fig 1, illustrate the time of convergence of it to the real evolution and the functional using (19) has the high level of convergence and is depicted in Fig 2:

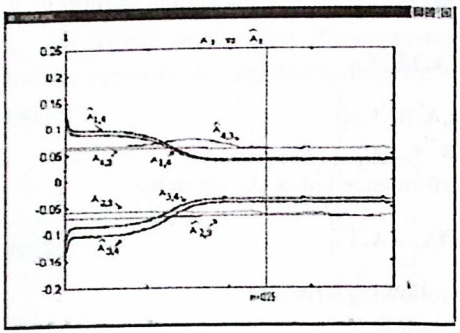


Fig 1. RTMADF (Parameter estimator)

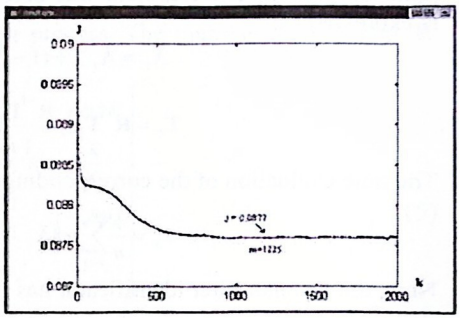


Fig 2. Time evaluation of J_n.

Finally, the simulation results by Kalman filter as RTMADF (was used the expressions (20) and (21)) has the high level of performance and has too the equal time of convergence with respect to the real system modeled by (14),

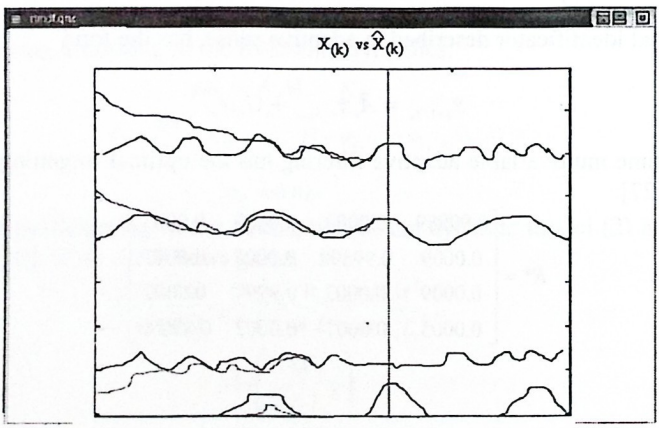


Fig 3. Time evaluation of $x_n - \hat{x}_{n \cdot Y_{n-1}}$.

These simulations showed that the RTMADF accomplished in the same interval the best results in a convergence sense considering in it the expressions (18) to (22).

5 CONCLUSIONS

In this work, we have exposed the basic concepts about the real-time multivariable adaptive digital filter (RTMADF).
The basic concepts considered were the properties about the real-time systems, and adaptive digital filters, both in multivariable sense; such that in agreement to [1], [7] and [3], required us the vector properties with respect to:
a) The number of vector operations required in the multivariable adaptive filter process by each iteration, bounded with time constraints respect to dynamical process conditions [8] and,
b) The filter tasks vector were scheduled, such that the adaptive filter could be implanted.
With respect to Multivariable Adaptive Digital Filter, we said that it was compound by two steps, described these as: the estimation and identification sections, developed in section 3, considering the Instrumental Variable and Kalman Filter by no stationary output system, respectively.
The RTMADF implementation was solved and either illustrated using a DC derivate multivariable novel model. In this case, we were requiring the dynamical arrays sets with message passing priorities, without to lose the synchrony between the whole of tasks vector.

REFERENCES

[1] Medel, J.J., Guevara, L. P. Comparación de la Dinámica en Tiempo Real de los Métodos Mínimos Cuadrados y Variable Instrumental para Estimación de Parámetros. Congreso Argentino de Ciencias de la Computación. Buenos Aires, Argentina. CACIC (2002).
[2] Chui, C., Chen, G. Kalman Filtering with Real-time Applications. USA. Springer, (1999).
[3] Gustafsson, F. Adaptive filtering and change detection. Sweden. Linkoping University John Wiley & Sons (2000).
[4] Papoulis, A., Bertran, M. Sistemas y circuitos. Spain, Ed. Marcombo (1989).
[5] Liu, J. Real-time Systems. USA. Prentice Hall (2001).
[6] Baras, J. Symbolic and numeric Real-time signal processing. USA. Technical Report University of Maryland (1999).
[7] Medel, J. J., Guevara L. P. Análisis restrictivo para filtros digitales en tiempo real presentando un ejemplo para un SLIT tipo SISO. Sevilla, Spain. Congreso Iberoamericano de Inteligencia Artificial IBERAMIA (2002).
[8] Nyquist, H. Certain Topics in Telegraph Transmission Theory. USA. AIEE Transactions (1928).
[9] Ash Robert D., Real Analysis and Probability, Academic Press (1972).

[10] Liu, C., Layland, J. Scheduling algorithms for multiprogramming in hard real-time environment. USA. Vol. 20, No. 4. ACM (1973).

[11] Ljung. L., and Gunnarson. T., Adaptation and tracking in system identification a survey, Automatica, 26(1), (1990) 7-22.

[12] Xue. P., and Liu. B., On asymptotic convergence of the least square algorithm, IEEE, International Symposium on Circuits and Systems, 1, (1991) 544-547.

[13] Guo. L., Ljung. L., and Prioret. T. Tracking performance analysis of the Forgetting Factor recursive least square algorithm, IEEE, Conference on Decision and Control, 1, (1992) 688-693.

[14] Tsakalis, S., and Limanond, S. Adaptive control of time-varying systems: An application to the attitude-momentum control of the space station, IEEE, Conference and Decision and control, 2, (1992) 1285-1286.

[15] Ward, D., Barren, R.,Carley, M. and Curtis, T. Real-time parameter identification for self-designing flight control, IEEE, National Aerospace and Electronics Conference (NAECON), 1, (1994) 526-531.

[16] Bittanti, S. and Campi, M., Bounded error identification of time-varying parameters by recursive least square techniques,. IEEE, Transaction on Automatic Control, AC 39, (1994) 1106-1110.

[17] Canetti, R. and Espana, M. Convergence analysis of the least-squares identification algorithm with a variable forgetting factor for time-varying linear systems, Automatica, 25 (1989) 609-612.

[18] Bittanti, S. and Campi, M. Least square identification of autoregressive models with time-varying parameters, IEEE, Conference on Decision and control, 4, 36 (1994) 3610-3611.

[19] Lindoff, B. and Holsts, J. Bias and covariance on the RLS-estimator with exponential forgetting in vector autoregressions, Technical report, Lund Institute of Technology, Department of Mathematical Statistics (1995).

[20] Distribution of RLS-estimator in time-varying AR(1)-process, in 5th IFAC Symposium on Adaptive Systems in Control and Signal Processing, Budapest, Hungary, (1995) 135-140.

[21] Parkum, J., Poulsen, N. and Holst, J. Recursive forgetting algorithms, International Journal of Control, 55, (1992) 109-128.

[22] Porat, B. Transactions on Acoustics, Speech and Signal, ASSP 33, (1995) 1209-1212.

[23] Bittanti, S., Bolzern, P. and Campi, M. Adaptive identification via prediction-error directional-forgetting factor: Convergence analysis, International Journal of Control, 50, (1989) 2407-2421.

[24] Bittanti, S., Bolzern, P. and Campi, M. Convergence and exponential convergence of identification algorithms with directional forgetting factor, Automatica. 26(5), (1990) 929-932.

[25] Mayne D. Q., A method for estimating discrete time transfer functions, Advances of Control, 2nd UKAC Control Convention (1967).

[26] Lindoff. B., and Holst, J., Convergence Analysis of the RLS identification algorithm with exponential forgetting factor in stationary ARX-structures, SIAM on Optimization and Control (1996).

[27] Poznyak A., and Medel J, J. J., Matrix Forgetting Factor, Int. Journal of Systems Science (1999).

[28] Gibson, J. E., and Tuteour, F. B. Control System Components, McGraw-Hill, NY, Toronto, London (1958).

[29] Rachid, A., On induction Motors Control, Control Systems Technology, IEEE., Control Systems Society, 5-3, (1997) 380-382.

[30] Medel J. J., and Poznyak, A. S. Adaptive Tracking for DC- derivate motor Based on Matrix Forgetting. México, Computación y sistemas (2001) 201-217.

[31] Medel J. J. "Análisis de dos métodos de estimación para sistemas lineales estacionarios e invariantes en el tiempo con perturbaciones correlacionadas con el estado observable del tipo: Una entrada una salida" Computación y sistemas volumen 6 numero 1, México (2002).

[32] Haykin S. "Adaptive filter theory". Prentice Hall information and system sciences series, USA (1991).

[33] Caines P. "Linear Stochastic Systems". Ed. Wiley, Canada (1986).

[34] Krten, R. Getting started with QNX 4, PARSE Software Devices, Canada (1998).

[35] Guevara, L. P., Medel, J. J. Implementación de un Filtro Digital en Tiempo Real sobre la plataforma QNX RTP 6.0. Conferencia de Ingeniería Eléctrica. CINVESTAV, México (2002).

[36] Guevara, L. P., Medel, J. J. Análisis restrictivo para filtros digitales en tiempo real presentando un ejemplo para un SLIT tipo SISO. Congreso Iberoamericano de Inteligencia Artificial IBERAMIA 2002, Sevilla España, noviembre (2002).

Steps Towards the Development of a Portable and Cost-Effective System for Human Resources Training in Small Satellite Technology

Esaú Vicente-Vivas[1], Israel Nava[2], Victor Lomas[2], Jorge Morales[2], E. A. Jiménez[3], Z. Carrizales[3] and F. Gómez[3]

[1]Instituto de Ingeniería, UNAM, Cd. Universitaria Coyoacan, 04510, México DF., evv@servidor.unam.mx ,
[2]Student, Posgrado FI-UNAM, Cd. Universitaria, Coyoacan, 04510, México DF., israel_nava75@hotmail.com, vlomasb@yahoo.com.mx, jmorales@iingen.unam.mx,
[3]Student, FI-UNAM, Cd. Universitaria, Coyoacan, 04510, México DF., factore8@hotmail.com, zairalilian@yahoo.com.mx, paco_84df@yahoo.com .

Abstract. A detailed global description of a cost-effective portable system for training purposes in the satellite technology field is presented. The system might be employed in both: school laboratories and Mexican research centers. The former would allow attracting young generations to the world of Science and Technology, and the latter would employ it to generate and validate new satellite technology. The preliminary development steps towards the construction of a demonstrative prototype are described. The whole system consists of a small, fully instrumented and Portable Didactic Satellite (PDS), operations software for the didactic satellite, as well as executable software for personal computers which performs PDS supervision and control. Full information about each one of the PDS subsystems, comments regarding its envisioned operations, and the current development process to accomplish the first demonstrative prototype are described in this paper.

Keywords: didactic satellite, portable laboratory equipment, training system, cost-effective technology.

1 Introduction

Mexico has developed in the past a couple of small satellites projects, and currently is developing a couple more of them. Those projects are very important not only to stretch the enormous technological gap with industrialized countries, but also to launch a domestic satellite development and research program. The last could take us in the medium and long term to generate our own Remote Sensing and Communications satellites, according to the needs of the country. However, to conduct such a line

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 117-130

of work, it is required the formation, qualification, and training of human resources in these technological areas.

According with our experience, it has been seen that the development of an experimental satellite in our country requires of 4 to 8 years of work, depending on the obtained financial support, as well as on the magnitude of the technological and scientific objectives.

We have also detected that this type of projects allows the participation of 25 to 200 people according to the challenges and the magnitude of the technological achievements faced by the satellite mission. Considering the cost of this type of projects (half million to 5 million US dollars) it is seen that the relation between the amount of participant personnel versus project cost is very low. In addition, if we consider how the young Mexican University people becomes attracted and motivated by satellite and space projects, it is observed an extremely low efficiency in the rows of participation and formation of new human resources in this field. By the way, this technology arena is related with the possibility to generate alternative solutions to national security problems from our country.

The analysis of the exposed scenario in México for the last 15 years, took us to elaborate the idea of developing a cost-effective training system for satellite technology employing commercial-off-the-shelf parts. The system is planned to be affordable enough to be used in laboratories with the intention to offer attractive, fast, and versatile training courses in the field. We plan to use the system in High Schools, Institutes, and Universities, with the intention of approaching new generations to the exciting world of space applications. Besides, the system will support the diffusion of Science and Technology activities in Mexico.

In order to carry out this project in a fast form, it was planned to be based on previous developments for satellite instrumentation elaborated by the Instituto de Ingeniería (IdeI) UNAM, [1], [2] and [3]. In this way, the paper describes the preliminary design of the PDS for human resources training in the satellite field. The PDS employs few satellite products already elaborated for the SATEX experimental microsatellite, [1], [2] and [3], as well as parts and systems of commercial use from the Automotive, Telecommunications, and Services industries.

1.1 Further Applications of PDS Computer and Related Work

With the intention to increase the applications of the equipment and subsystems generated by our group, we planned to develop a small sized single board computer to be used not only in the PDS, but also in 1 kg picosatellites (PICOSATs) like the one projected to be developed jointly by IdeI UNAM and IPN.

1.1.1 PICOSATs Around the World

The PICOSAT concept was developed by Stanford University and California Polytechnic State University in San Luis Obispo under the name of CubeSat in 1999, [4]. Since then a great amount of universities from all over the world are developing, are planning to build, or have developed a PICOSAT project to enter the space activities. Among the countries involved in CubeSats projects are: USA, Canada, Germany, Japan, Denmark, Netherlands, Norway, Switzerland, Australia, Korea, Malasya, Argentina and Colombia, [5]. The cost of this type of projects depends on the country that makes it, and go from <100,000 to 1,500,000 US dollars including the costs of inception, launch, operation and end-of-life.

PICOSATs are launched in groups as secondary payloads on rockets from different countries. However, Russian rockets are often selected for launching purposes because of its attractive cost. Even the Space Shuttle has opportunities to launch important groups of them. In 2006 the launching cost for orbiting a PICOSAT with a group of CubeSats through a Russian rocket was 40,000 US dollars. There are a couple of companies in USA that commercialize the fundamental subsystems to construct PICOSATs. They charge 6,000 US dollars for every subsystem such as: structure, communications, electrical power, attitude determination & control, and on board computer. Nevertheless, it is important to notice out that additional investment has to be considered for experiments, development of operations software, hardware environment testing, launching services and ground station. In this sense, the cost required to develop not only a PICOSAT but even the acquisition cost of every PICOSAT commercial subsystem is still a problem for many universities from developing countries. Besides it has to be taken into account that low-earth-orbit CubeSat missions have a typical lifespan of 3-9 months. In other words, for a University from a developing country means an important investment to place a small spacecraft in orbit for a very short period of time.

Another important point to highlight refers to the widespread trend in the PICOSAT field to use commercial-off-the-shelf (COTS) electronics parts, must of them even without protections for space operation. For this purpose few universities perform laboratory testing on parts being subjected to heavy ion testing, vacuum, and temperature extremes to select those parts with inherent capabilities to survive harsh environments, [6]. In order to determine the expected amount of radiation in space, specialized Radiation simulation software from the US navy is available through the web, [7]. When COTS parts are taken to space orbit and when they behave well, then they become referred to as space proven or flight tested. When those parts are published in papers other PICOSAT builders with less resources take advantage of that information and decide to employ them (power electronics, computers, communications, sensors, solar cells, etc) in their small satellites. Some times PICOSAT parts are successfully tested in space even without previous testing in laboratory, those parts are also referred as space proven or flight tested.

CubeSat missions are attractive and demanded around the world because they are employed for technology demonstrations, proof-of-concepts, scientific experiments and human resources training, [8], [9], [10], and [11]. However, because of financial problems most of the Universities from developing countries are impeded to participate in this strategic technology field. This scenario along with our experience in the small satellite field took us to develop the PDS with the goal to provide Universities from our country with a portable and cost-effective system for human resources training in the small satellite field.

On the other hand, the development of a PICOSAT has extremely severe restrictions in several issues, one of them related with the maximum dimensions every subsystem can reach. For these reasons the size of the PDS computer was projected to fulfill the size restrictions of a PICOSAT. Later, the size of the PDS computer defined the dimensions of the complementary PDS subsystems. It should be highlighted that the PDS is not intended for space use but for educational purposes in terrestrial laboratories.

1.2 Availability of Commercial Educative Satellites in the International Market

It must be stated that commercial availability of similar products to the PDS is rarely seen in the global market. Right now the only PDS like commercial product detected by the authors is the Eyasat educational system developed initially by the US Air Force and since February 2006 commercialized by Colorado Satellite Services, [12]. The Eyasat basic equipment starts at 8,000 dollars. However, this price is difficult to be afforded by Mexican schools, and in general difficult to be afforded in developing countries. This is why the PDS is planned to be developed and offered for an under 3,000 US dollars cost in order to be attractive for different domestic schools, universities, and so on. This goal shaped the design and the main characteristics of the PDS in order to achieve a cost-effective development. To accomplish the referred objective all the PDS PCBs will be manufactured in only two layers, the power subsystem was achieved with rechargeable batteries and an external battery charger, 1-wire sensors were preferred, and the PDS structure was chosen according with cost-effective commercial available products. In addition, few cost-effective auxiliary interfaces such as the command execution visualization subsystem, orientation determination sensors, serial and USB expansion ports, as well as expansion cards for user defined interfaces are aggregated in the PDS to be perceived as a friendly device with growing capabilities.

2. Portable System for Human Resources Training in the Satellite Technology Field (POSTSAT)

The POSTSAT is basically formed by the PDS, the PDS operations software, and executable software for personal computers. The PDS will contain an operations control software to carry out, among other functions: 1) The communication with the personal computer software, that will perform as the Ground Station Software (GSS)

for Telemetry acquisition and Command Shipment, 2) the functions requested by the GSS, 3) the capture of requested telemetry, 4) the pack up of telemetry to be transmitted by a wireless mean, 5) the manage of GSS required protocols, 6) the simulation of orbital times for the didactic satellite, 7) real time operations with the didactic satellite, 8) the stabilization functions, etc.

As shown in figure 1, the POSTSAT contains a PDS with a cylindrical shape of 12 cm in diameter and 17 cm as maximum height. In addition, it contains the Ground Station Software (GSS) that executes itself in a PC. This carries out the functions of telemetry acquisition and command transmission to the PDS. Besides, this software allows uploading new operation software to the PDS by wireless means.

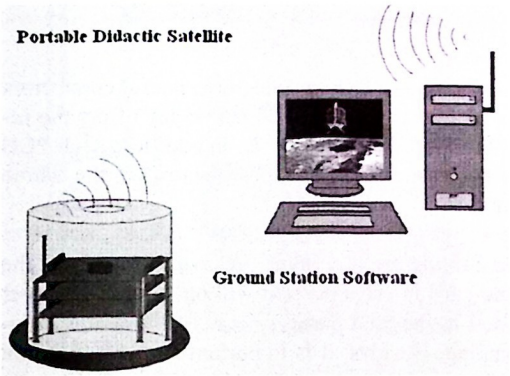


Figure 1. Fundamental parts of the POSTSAT.

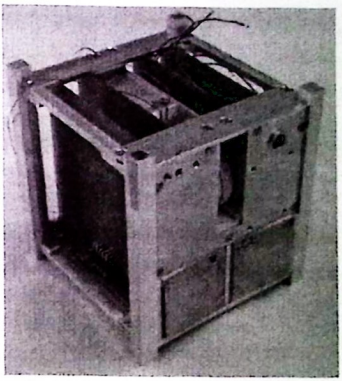


Figure 2. One Kg mass PICOSAT

The whole POSTSAT software will be recorded in a compact disk (CD) which will also contain the user and the operations manuals. For these reasons, the POSTSAT final version will consist only of the PDS and a CD. This is the system we expect to share with educative institutions from our country.

In addition to the system that is described in this article, it will be elaborated later a laboratory practice manual. This will allow carrying out training practices with the satellite system in a simple and friendly way, in order to teach the satellite system operation to the user. This manual will be carefully elaborated by another work team. It will emphasize the clarity of concepts, as well as the progressive transference of information, besides of reinforcing the learned knowledge, the questioning of treated subjects, as well as the comparison of learned concepts with the operation of commercial small satellite systems. The organization, development, and implantation of this learning and training methodology will appear in later publications.

3. Architecture of the Laboratory Didactic Satellite

The PDS architecture is basically defined by its computer dimensions, which were fixed according with the dimensions of a PICOSAT, [11], [13], and [14], figure 2. The PICOSAT is a 10cm x 10cm x 10cm aluminum frame cube, which admits the installation of 8.5cm x 8.5cm printed circuits boards (PCBs). In this way, the same PCB size was fixed for the PDS electronic cards, whereas its complementary subsystems will reside in printed boards of the same size in a tandem assembling, figure 1. The complementary PDS subsystems are: Power, communications, sensors, inertial wheel stabilization, magnetic torquer coil stabilization, visualization of satellite answers to commands, and a card for new developments.

3.1 Computing subsystem

The PDS has a single board operations computer, which integrates lateral connectors in a bus way to interconnect cards in tandem. The electrical connector offers the required mechanical subjection among the printed circuit boards. In addition, each PCB contains screw holes in each one of the corners, which allows screwing the whole printed board array to the PDS structure.

When using bus type connectors, all the electrical signals are available for all the assembled cards, thus it is possible the card interconnection without caring the order of them. Therefore, the order mentioned in the next paragraphs is just with the intention to have an order in the article writing. Besides, it is important to notice out that all the bus type connectors in the cards are male type in their superior part, whereas the inferior part looks like a wire-wrap connector. This allows the interconnection of cards in tandem as much by the superior side as by the inferior.

Considering that the PDS computer will be used as well in Picosatellites to be placed in space orbit, the computer design included hardware to eliminate the latch-up effect. This effect is generated by cosmic radiation in commercial electronic parts when they are taken to space orbit. The latch-up prevention hardware will be installed exclusively in computers elaborated for PICOSATs, whereas the PDS computers will not install such electronic parts because they will be employed only in terrestrial environment. The goal is to generate a single PCB computer for both applications to increase the applications of this computer, as well as lowering the development cost of our small satellites projects. Figure 3 shows the block diagram for the PDS computer.

It should be highlighted that the remaining cards exposed in following paragraphs will be used only in the PDS and are not intended for use in space orbit. However it is also important to mention that the world-wide philosophy to develop PICOSATS is based on the extensive use of commercial-off-the-shelf parts and systems. Where must of the projects even excludes the use of latch-up protection electronics. In this way PICOSATS have attractive development times (six months in some cases) which make them useful to validate state-of-the-art technology in space. In addition, they can be manufactured with very small budgets. This philosophy has taken the

PICOSAT field to the Faster, Cheaper and Better approach. As can be deduced, the expected life for this type of spacecraft is low but at the same time enough to test and validate new technologies in a short period of time.

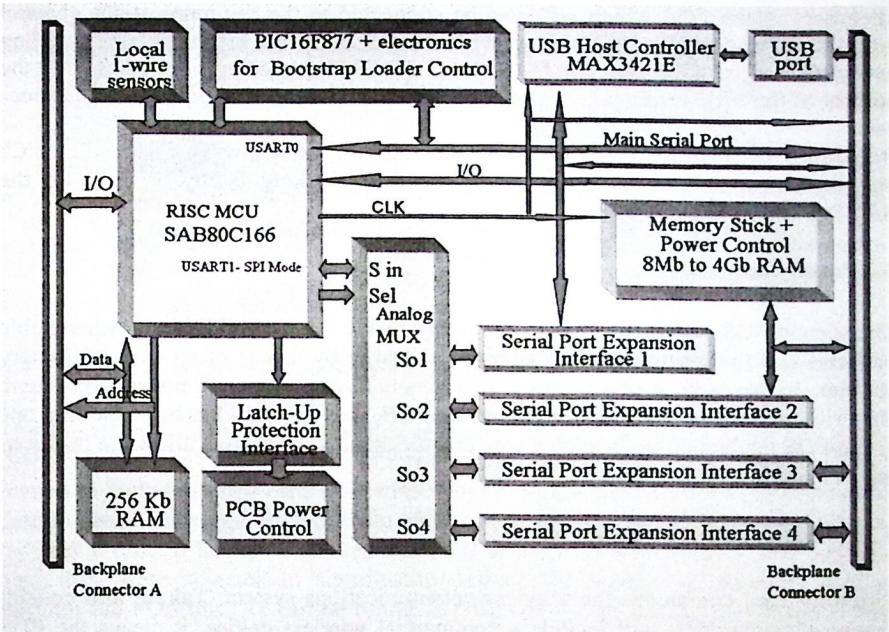


Figure 3. Preliminary block diagram for the PDS computer.

The PDS single board computer is built around the 16 bit RISC SAB80C166 processor from Siemens, industrial version with extended temperature, 40 Mhz oscillator, 256 kb of RAM memory where the PDS operations program will be loaded, hardware for automatic uploading of new programs to the computer and a total of 5 serial ports. The lasts support full-duplex asynchronous communication up to 625 Kbaud and half-duplex synchronous communication up to 2.5 Mbaud. The synchronous mode is employed in serial port So1 to gain access to a USB port on one side, and on the other side serial port So2 allows the communication with a Memory Stick card, [15]. The last will be used to add as much memory as required in order to store several types of PDS telemetry. It should be highlighted that the design of the computer hardware is already finished, however software development is required to validate the communications with both the USB port and the Memory Stick interface.

On the other hand, the 100-pin SAB80C166 microcontroller internally contains important resources as follows: a watch dog timer, interrupt controller, 16-bit timers, 10-channel 10-bit A/D converters, two serial channels and several 16 bits I/O ports, with a total of 76 I/O lines.

The SAB80C166 allows the uploading of new programs in external RAM memory, however it is required to control few electrical signals according with specific feedback responses from the processor. For this purpose a small PIC microcontroller was integrated in the PCB which will also be connected to the communications channel with the Ground Station. In this way, when the GSS software sends the "downloading new program" command to the PDS, the PIC16F877 microcontroller takes over the control of the SAB processor for both to achieve and to supervise the uploading process.
Regarding the software development for the SAB80C166 it is written in standard "C" language, programs are compiled using the BSO Tasking family of tools for the SAB80C166 microcontroller.

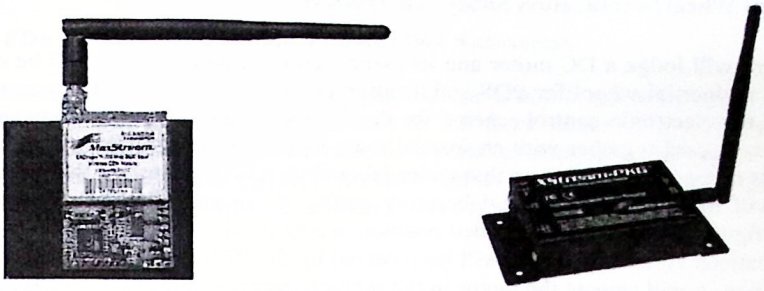
3.2 Power Subsystem

The second PDS card is the power one, where reside 4 AA-sized lithium rechargeable batteries and the electronics that admit its recharge by means of an external battery charger. In this way, it constitutes a simple, small, and economic power subsystem. Batteries are located in the lower part of the PCB, thus this card has to be the first one in the PDS tandem array, in such a way that mechanical separators allow the positioning of batteries.

3.3 Communications Subsystem

The third card constitutes the wireless communications system. Taking into consideration that this PCB will include a commercial wireless device, it means the PDS will raise its cost when it includes the wireless system, since its intention is the PDS to become a commercial system. However, this subsystem confers a great operative versatility to the didactic satellite. We are considering the integration of a Stand-Alone RF radio modem from MaxStream, Inc., particularly the XStream-PKG-R RF Modem, operating in the 2.4 GHz band, indoor range of 180 m, 2.1 dB dipole antenna, throughput data rate from 9,600 to 19,200 bps, serial communications, 7 to 12 V operation, spread spectrum, FSK modulation, 7x14x3 cm size and 200g weight, see figure 4.

It is necessary to point out that the PDS can be used also without the wireless communication system, in which case, this PCB holds only an RS232 line driver to accomplish the RS232 voltages for a PC serial port. Under this configuration, the communications channel among the educative satellite and the PC is just a wire. In this way, the PDS has two operative options from the communications point of view, which will significantly influence the cost of the didactic satellite.



(a) (b)

Figure 4. RF radio modem being considered for the project, a) For a PDS PCB and, b) For ground station.

3.4 Platform and Orientation Sensors Subsystem

The fourth card integrates the satellite telemetry sensors. Although, each one of the described cards has at least a pair of local sensors (current and temperature), we have planed to include in this card a tilt sensor and a digital compass for PDS orientation determination purposes. Both of them will be very useful to visualize the PDS telemetry in real time. It has to be noticed that most of the PDS sensors are 1-wire technology, they were selected to significantly reduce the number of traveling signals through the PDS bus. As can be found in [16], the 1-wire sensors let the microcontroller, by means of only two wires, to access an important amount of sensors. This contrasts with conventional sensors, which require two wires for each included sensor.

Regarding the tilt sensor and the digital compass, will be used in the PDS under a real time telemetry acquisition mode requested by command, which will allow the GSS software to generate 3D real time animations in connection with the satellite manipulation generated by the user in laboratory environment.

The PDS will employ the 1490 digital compass from the Robson Company. It is a solid-state Hall effect cheap device, 12 pin component, with cylindrical shape and requires 5 V supply, [17]. When rotated it senses the position of the four cardinal points on a compass, North, South, East and West. As well as the intermediate directions: North East, North West, South East, and South West.

For the tilt sensor a low cost SQ-SI-360DA solid-state MEMS based Inclinometer which provides both an analog voltage output and digital serial output corresponding directly to a full-scale range of 360° of pitch angle (single axis range) or ± 80° of pitch and roll angle with (dual axis range) will be employed, [18]. The device offers 0.1^0 resolution when the digital serial output is used. On the other hand, when analog output is utilized a 2° resolution is obtained.

3.5 Inertial Wheel Stabilization Subsystem (IWSS)

A fifth card will lodge a DC motor and its associated electronics, which will be employed as an inertial wheel for PDS stabilization control effects. We are currently validating the electronic control scheme for the DC motor and making arrangements to generate a card together with an inertial mass located in the rotation axis of the motor. It is necessary to point out that under laboratory validation testing the didactic satellite will be suspended in the laboratory ceiling by means of three strings, as shown in figure 5. Once in a suspended position, a command will be sent through the communications system, and then will be received by the PDS computer, then, as an answer action, it will actuate the motor in the rotation direction and at the RPMs indicated by the command. Under these circumstances, the didactic satellite will experience reaction forces that will generate the PDS movement. This will allow the user to carry out several dynamic experiments for satellite stabilization control in a cost-effective fashion.

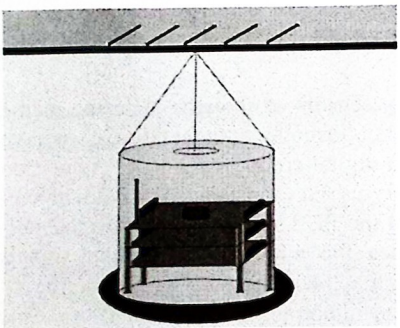


Figure 5. PDS suspended from the ceiling. Stabilization testing procedures.

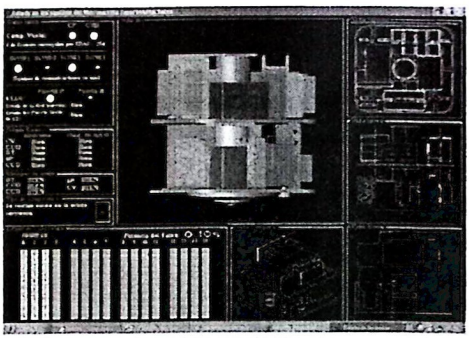


Figure 6. Ground station software for the PDS will be inherited Satex microsatellite project.

The IWSS operation will be autonomous from the PDS computing system to allow the continuous control and supervision of the PDS stabilization subsystem. For this purpose a PIC processor 18F4431 will be integrated in this board, while the Microchip MPLAB C18 C compiler will be employed for software development. The communication among PDS computing system and PIC microcontroller will be made through the serial port expansion interface 3 shown in figure 3.

3.6 Stabilization Subsystem based on Magnetic Torquer Coils

The PDS has also been planned to include an electronic card lodging a number of small magnetic torquer coils (MTC) that will generate small forces to affect its control dynamics. These will be operated in diverse modalities with the momentum wheel interface to implement different stabilization procedures with the satellite training system.

3.7 PDS Command Execution Visualization Subsystem

The seventh card will integrate a display device which is currently under definition stage. This will provide visual feedback to the users whenever the PDS receives a new command from the GSS executed in a personal computer. Among the interface possibilities hexadecimal displays, led stack lights, sound interface and a Digi-Talker interface (to provide talking capabilities to the PDS) are found.

3.8 New Developments Subsystem

The PDS will offer one or few expansion boards (as required) with access to the resources (computer, power, communications and command visualization) of the satellite training system. This is achieved by the PDS back plane type connectors provided at any one of the development cards. In addition, the card will have a proto-board layout to ease the placement of passive and active components, as well as to allow wire-wrap connections.

3.9 PDS Structure Subsystem

In order to define the didactic satellite structure two different choices were considered. The first one was constituted by a rectangular prism shaped acrylic structure, and a second one conformed by a cheap commercially available cylindrical container manufactured with plastic. When it was considered the manufacture cost, the manufacture time and the acquisition of materials, it was decided to go for the commercial cylindrical container. In addition, the last solution allows the POSTSAT to be produced at a lower cost through the reduction of manual labor. The chosen structure is light, cost-effective and will shape the PDS to render a satellite appearance. Once a set of PCBs are plugged together, the stack will be fixed to the plastic structure with screws.

4 Ground Station Software for Remote PDS Control and Remote PDS Supervision

The GSS software executes in personal computers and interacts with the PDS software to achieve both the demonstration and the training functions of the didactic satellite system. In fact, great part of the friendship attributes for the whole POSTSAT system (simplicity and clarity of use) are generated by this software. For these reasons, it was decided to develop the software under the bases of a previously developed software for the Satex microsatelite project [3]. This software performs similar functions to those planned for the PDS, we have access to the source code, and importantly, it was tested during the development stages of the referred project. The software written in Visual Basic employs enough visual resources so that the user may easily identify the satellite status from telemetry data. The visual resources are

generated from AutoCad drawings and the animations are rendered with Flash Software, figure 6.

Despite GSS software will be based on the Ground Stations software developed for the Satex project, it will be a modified version according with particular characteristics of the satellite training system. However, the software will have similar functions to those employed to control small satellites launched into space orbit, among them: 1) Immediate answer commands, 2) Task commands to define assignments according with programmed times defined by the command itself and which will have delayed answers from the PDS, 3) Telemetry request commands in agreement with time intervals specified by the user, 4) Real time telemetry requests to allow on-line PDS supervision with the help of 3D animations, 5) Time programming (scaled) required by the PDS to simulate orbital timings in agreement with user requirements, 6) Satellite stabilization control commands, etc.

4.1 Communications Subsystem for the PC

As already mentioned, the communication between GSS and the PDS software will be allowed either by cable or by wireless means, so that the POSTSAT system will be used in agreement with financial resources available in many laboratories. Figure 4b shows the equipment being considered for these purposes. It will be directly connected to the serial port of the personal computer where the GSS will be executed.

5 Current State of the Project and Planned Activities

So far we have important advances in every PDS subsystem that is why we expect to assemble a preliminary prototype and get it available this year. However, following activities will be required to be developed in parallel fashion to reach the goals of having a full demonstrative and useful system: 1) development of PDS operative software, 2) development of ground station software, and 3) Elaboration of manuals for the whole system. It should be highlighted that the success of the exposed system will also be indeed in the planning, in the methodology and in the content of the system manuals.

Once the first complete and functional prototype is finished, few copies will be assembled to use them in a pilot training program in the Engineering Postgraduate course titled "Design and construction of small satellites", given by the first author at the Posgrado de Ingeniería, UNAM, since year 2004. With the feedback of the participant students and with the results of use and evaluation, the necessary adjustments to the manuals and program course will be made. Then a diffusion stage will be started to incorporate the POSTSAT system to specific programs from Schools interested in the exposed system.

6 Concluding Remarks

We have presented the preliminary design, the development advances, the global architecture and the projected operations for a portable and cost-effective system to train human resources in the field of small satellite technology. The equipment was planned and based on 15 years of experience and work in the area. The PDS also considers the attraction of young people from our country for the satellite and the space fields. On the other hand, we expect the project will be affordable for many Institutes, Technologic Schools and Universities, and hence will represent an open gate for the new generations to participate in a field of work perceived to be far away from the academic possibilities of developing countries. In addition, the exposed system will attract young people to the world of Science and the Technology.

In this way, the exposed project tries to generate a whole portable training system, attractive and friendly, with capabilities to be adopted as a partner to access the world of science and technology, satellite technology or towards other technological fields such as Electronics, Telecommunications and Informatics.

7 References

1. Esaú Vicente-Vivas, Fabián García-Nocetti and Francisco Mendieta-Jiménez, "Automatic maintenance payload on board of a Mexican LEO microsatellite", Acta Astronautica Journal, Elsevier Science,, Volume 58, Issue 3, Pages 149-167, February 2006.
2. E. Vicente-Vivas y D.F. García Nocetti, "Computadora de Vuelo Triplex de Diseño y Manufactura Mexicana para el Microsatélite Satex", Revista Iberoamericana Información Tecnológica, ISSN 0716-8756, Vol. 17, No 1, pp. 69-76, Enero 2006, Chile.
3. E.Vicente Vivas, A.Espinoza M., C. Pineda F., J.R. Tórres F. Y A. Calvillo, "Software de Adquisición de telemetría y control de operaciones para microsatélites", 4° Conferencia Internacional en Control, Instrumentacion virtual y Sistemas Digitales "CICINDI 2002", Pachuca, Hidalgo, México, Agosto del 2002.
4. CubeSat Origins, http://directory.eoportal.org/pres_CubeSatConcept.html .
5. CubeSats Running Projects, http://en.wikipedia.org/wiki/CubeSat#Current_ running_projects .
6. Y.Tsuda et al, "University of Tokyo's CubeSat "XI" as a Student-Built Educa tional Pico-Satellite -Final Design and Operation Plan", The 23rd International Symposium of Space Technology and Science, Matsue, Japan, 2002.
7. Nav Cosmic Ray Effects on Micro Electronics, https://creme96.nrl.navy.mil/ .

8. The Stensat Group, http://www.stensat.org/ .
9. H. Heidt et al, "CubeSat: A new Generation of Picosatellite for Education and Industry Low-Cost Space Experimentation ", 14th Annual/USU Conference on Small Satellites, August 2001.
10. Montana Earth Orbiting Pico-Explorer, http://www.ssel.montana.edu/merope/ .
11. Miyamoto K. et al, "Tokyo Institute of Technology Small Satellite Projects", 7th University Space Systems Symposium, Waikoloa, Hawaii, November (2004).
12. Eyasat home page: http://eyassat.com/tiki/tiki-index.php?page=ContactUs .
13. The Stanford University CubeSat concept, http://cubesat.info/ .
14. CUBESATS - CanX-1 PicoSat – at University of Toronto, http://www.utias-sfl.net/nanosatellites/CanXProgram.html .
15. Memory Stick Standard, Memory Stick PRO Specification Summary, Sony Corporation, December 2004.
16. Maxim, http://www.maxim-ic.com/1-Wire.cfm, (2005).
17. Digital Compass from Robson Company, http://www.imagesco.com/articles/1490/03.html .
18. Inclinometer%20Datasheet_3.4c.pdf, http://www.signalquest.com/SQ-SI-360DA_ Solid-State%20MEMS %20Analog%20and%20Digital%20

Teaching the Design of Takagi-Sugeno Fuzzy Controllers for an Under-Graduate Course

R. Rajesh[1], T. Sureshkumar[1], M. R. Kaimal[2], and K.K. Suresh[3]

[1] Department of Computer Science and Engineering, Bharathiar University, Coimbatore - 641046 India
kollamrajeshr@ieee.org, btsuresh2@yahoo.com
[2] Department of Computer Science, University of Kerala, Thiruvananthapuram - 695581 India
mrkaimal@yahoo.com
[3] Department of Statistics, Bharathiar University, Coimbatore - 641046, India

Abstract. Fuzzy control can be used to improve existing controller systems by adding an extra layer of intelligence to the current control method. This paper deals with a survey of Takagi-Sugeno fuzzy model and its controller. Two problems namely, inverted pendulum model and ball-and-beam model have been simulated and the results are shown. This survey is used to teach the design of Takagi-Sugeno fuzzy controllers for under-graduate courses.

1 Introduction

Substantial changes have undergone in the past years in the curricula for computer science and engineering [2]. Industry also demands the changes in the curricula by introducing new courses with hands on experience in writing softwares for real applications. The combination of hands-on experience and computer simulation with the more traditional, theoretical lecture material provides a well-rounded learning experience that better prepares the students for implementation in the real world [4].

The traditional method of introducing undergraduates to the areas that involve fuzzy logic and fuzzy controllers is through upper level elective courses [1]. The curricula for such an elective should be designed in such a way that math/theory side should not be sacrificed in favor of demos using high-level packages, but at the same time keeping in mind, that students cannot implement even a toy system if they must do it all from the scratch.

With these thoughts in mind, this paper focuses on teaching the design of the newest technology of Takagi-Sugeno fuzzy controllers for undergraduate courses with little in-depth treatise on the theory of fuzzy sets and fuzzy control.

The required background for the students includes knowledge in fuzzy logic and good programming skill which is essential to develop and test various real time applications. MATLAB [3, 6, 19, 20] software is preferred for programming, since it allows students to investigate the characteristics of the algorithm and

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 131-140

easily design their algorithm with its vast assortment of graphical and simulation functions.

This paper is organized as follows. In section 2 a survey of TS fuzzy model and its corresponding PDC controller design techniques are discussed. Section 3 discusses the application of TS fuzzy controller for inverted pendulum system and ball and beam system. Section 4 concludes the paper.

2 Takagi-Sugeno fuzzy model and its controller

Modeling techniques are believed to contain hidden intelligence that is able to solve a problem whenever the experimenter doesn't know what he is doing. A model is just a mapping from a given input to a given output. The essential element for the study of a nonlinear control problem is to get a tractable model of a dynamical system for use in control system design. The design model should be simple enough to work with, but must retain the essential features of the process.

Literature says that Takagi-Sugeno (TS) fuzzy model[21] along with Parallel Distributed Compensation (PDC) [23] scheme is best for control system design[3],[7]-[19]. This section deals with the survey of TS fuzzy model and its controller.

One can construct a Takagi-Sugeno fuzzy model if local description of the dynamical plant to be controlled, given by,

$$\dot{X} = f(X, u)$$

is available in terms of linear models

$$\dot{x}(t) = A_i x(t) + B_i u(t), \quad i = 1, 2, \ldots, r$$

where the state vector $x(t) \in \Re^n$, the control input $u(t) \in \Re^m$, and the matrices A_i and B_i are of appropriate dimensions. The above information is then fussed with the available IF-THEN rules where the i^{th} rule can have the form

$$R_i: \text{ IF } x_1(t) \text{ is } F_1^i \text{ AND} \ldots \text{AND } x_n(t) \text{ is } F_n^i \text{ THEN } \dot{x}(t) = A_i x(t) + B_i u(t)$$

where F_j^i, $j = 1, 2, \ldots, n$, is the the j^{th} fuzzy set of the i^{th} rule. Let $\mu_j^i(x_j(t))$ be the membership function of the fuzzy set F_j^i and let

$$w_i(t) = \prod_{j=1}^{n} \mu_j^i(x_j(t)), \; i = 1, 2, \ldots r$$

where r is the number of IF-THEN rules. Then, given a pair (x(t), u(t)), the resulting fuzzy system model is inferred as the weighted average of the local models and has the form

$$\dot{x}(t) = \frac{\sum_{i=1}^{r} w_i(t)(A_i x(t) + B_i u(t))}{\sum_{i=1}^{r} w_i(t)} = \sum_{i=1}^{r} \alpha_i(t)(A_i x(t) + B_i u(t)) \qquad (1)$$

where

$$\alpha_i(t) = \frac{w_i(t)}{\sum_{i=1}^{r} w_i(t)} \text{ for } i = 1, 2, \ldots, r$$

Note that $\alpha_i(t) \geq 0$ and $\sum_{i=1}^{r} \alpha_i(t) = 1$.

The open loop system corresponding to (1) is

$$\dot{x} = \sum_{i=1}^{r} \alpha_i(t)(x) A_i x(t) \tag{2}$$

Theorem 1. *[22] The equilibrium of a fuzzy system (2) is asymptotically stable in the large if there exists a common positive definite matrix P such that*

$$A_i^T P + P A_i < 0 \quad i = 1, 2, \cdots, r \tag{3}$$

It is obvious that a necessary condition for the existence of a common symmetric positive definite P satisfying (3) is that each A_i be asymptotically stable; that is, the eigenvalues of each A_i be in the open left-hand complex plane.

Parallel Distributed Compensation (PDC) [23] can be utilized to design fuzzy controller rules. The idea is to design a compensator for each rule of the fuzzy model. The i^{th} rule of the controller is of the form

$$R_i : \text{ IF } x_1(t) \textit{ is } F_1^i \text{ AND} \ldots \text{AND } x_n(t) \textit{ is } F_n^i \text{ THEN } u(t) = -K_i x(t)$$

The overall fuzzy controller is obtained by fuzzy blending [23] of each individual linear controllers and is given by

$$u(t) = \frac{\sum_{j=1}^{r} -w_j(K_j x(t))}{\sum_{j=1}^{r} w_j} \tag{4}$$

which is nonlinear.

Substituting (4) into the fuzzy system (1), we obtain

$$\dot{x} = \sum_{i=1}^{r} \sum_{j=1}^{r} \alpha_j(x) \alpha_i(x) \{A_i - B_i K_j\} x \tag{5}$$

Theorem 2. *[24] The equilibrium of a fuzzy control system (5) [i.e., Fuzzy controller (4)+ Fuzzy Model (1)] is asymptotically stable in the large if there exists a common positive definite matrix P such that the following conditions are satisfied:*

$$\{A_i - B_i K_j\}^T P + P\{A_i - B_i K_j\} < 0 \tag{6}$$
$$i, j = 1, 2, \cdots, r$$

It is obvious that a necessary condition for the existence of a common symmetric positive definite P satisfying (6) is that each $A_i - B_i K_j$ be asymptotically stable; that is, the eigenvalues of each $A_i - B_i K_j$ be in the open left-hand complex plane.

The system (5) can also be written as

$$\dot{x} = \sum_{i=1}^{r} \alpha_i(x)\alpha_i(x) \{A_i - B_i K_i\} x \\ +2 \sum_{i<j} \alpha_i(x)\alpha_j(x) G_{ij} x \tag{7}$$

where

$$G_{ij} = \frac{\{A_i - B_i K_j\} + \{A_j - B_j K_i\}}{2} \quad i < j \tag{8}$$

Therefore the following sufficient condition is obtained.

Theorem 3. *[24] The equilibrium of a fuzzy control system (5) [i.e., Fuzzy controller (4)+ Fuzzy Model (1)] is asymptotically stable in the large if there exists a common positive definite matrix P such that the following two conditions are satisfied:*

$$\{A_i - B_i K_i\}^T P + P\{A_i - B_i K_i\} < 0 \tag{9}$$
$$i = 1, 2, \cdots, r$$
$$G_{ij}^T P + P G_{ij} < 0 \tag{10}$$
$$i < j \leq r$$

Pole placement method or linear quadratic control design method or Genetic algorithms or Linear Matrix Inequalities (LMI)[5, 22] can be used to find out the values of the Gains, $K_i s$. Since the course if for undergraduate students, the gain values are obtained using linear quadratic method without going in-depth treatise on the theoretical part or stability issues.

3 Control Applications

3.1 Inverted Pendulum

A famous benchmark problem namely the inverted pendulum control problem is chosen for study. The non-linear and non-stable behavior of the inverted pendulum problem renders the use of conventional method very difficult.

The inverted pendulum system consists of a pole hinged on a cart as shown in Figure 1a. The control objective is to balance the inverted pendulum for the approximate range of vertical angle, namely $x_1 \in (-\pi/2, \pi/2)$. The equations of motion of the pendulum [23] are

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = \frac{g \sin(x_1) - aml x_2^2 \sin(2x_2)/2 - a \cos(x_1) u}{4l/3 - aml \cos^2(x_1)}$$

where x_1 denotes the angle (in radians) of the pendulum from the vertical, and x_2 is the angular velocity, g is the gravity constant, m is the mass of the pendulum,

M is the mass of the cart, 2l is the length of the pendulum, u is the force applied to the cart and a =1/(m+M). In our simulation we used the values, g=9.8 m/s^2, m=2Kg, M=8Kg, a=0.1, 2l=1m, $\beta = \cos 88°$.

We approximate the system by the following two-rule Takagi Sugeno fuzzy model

$$R_1 : \text{ If } x_1 \text{ is about 0 then } \dot{x} = A_1 x + B_1 u$$
$$R_2 : \text{ If } x_1 \text{ is about } \pm \pi/2 \text{ then } \dot{x} = A_2 x + B_2 u$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \dot{x} = \begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} \quad A_1 = \begin{bmatrix} 0 & 1 \\ \frac{g}{4l/3 - aml} & 0 \end{bmatrix} \quad B_1 = \begin{bmatrix} 0 \\ \frac{-a}{4l/3 - aml} \end{bmatrix}$$
$$A_2 = \begin{bmatrix} 0 & 1 \\ \frac{2g}{\pi(4l/3 - aml\beta^2)} & 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 \\ \frac{-a\beta}{4l/3 - aml\beta^2} \end{bmatrix}$$

By utilizing the concept of Parallel Distributed Compensation (PDC), the following two rules are designed for the controller.

$$R_1 : \text{ If } x_1 \text{ is about 0 then } u_1 = K_1 x$$
$$R_2 : \text{ If } x_1 \text{ is about } \pm \pi/2 \text{ then } u_2 = K_2 x$$

By choosing the eigen values [-2+2i, -2-2i], the controller gains K_1 and K_2 are obtained as [-143.3333 -22.6667] and [-3315.6 -764] respectively. Figure 1b shows the plot of θ starting from 16, 30, 79 and 87 degrees. Figure 1c and 1d shows the plot of $\dot{\theta}$ and u respectively.

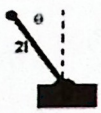
3.2 Ball and Beam System

A ball is placed on a beam, where it is allowed to roll along the length of the beam. A motor at the center of the beam will apply torque to the beam. See Figure 2a. When the angle of the beam is changed from the vertical position due to the torque, gravity causes the ball to roll along the beam. A controller will be designed for this system so that the ball's position can be manipulated. We assume that the ball rolls without slipping and friction between the beam and ball is negligible.

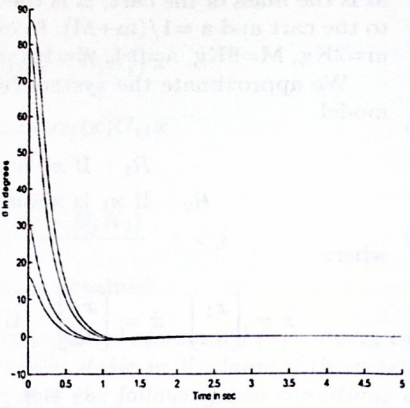
The equation of motion for the ball is given by

$$\ddot{r} = \frac{-mg \sin \alpha + mr\dot{\alpha}^2}{\frac{J}{R^2} + m} \tag{11}$$

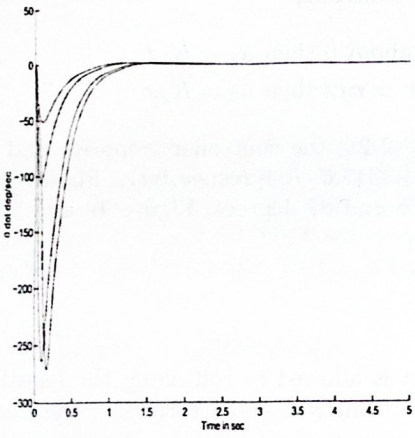
where, m is the mass of the ball, r is the radius of the ball, l is the length of the beam, J is the ball's moment of inertia, r is the ball position coordinate, α is the beam angle coordinate. In our simulation the values used are m=0.11Kg, r=0.015m, g=9.8m/s^2, J=9.99×10^6kgm^2.



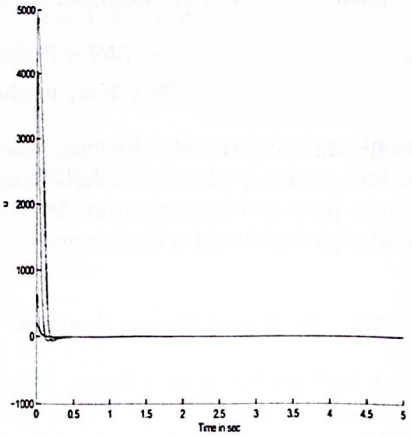
(a) Inverted pendulum



(b) Response of the θ



(c) Response of $d\theta/dt$



(d) Response of u

Fig. 1. Simulation results of Inverted Pendulum.

We approximate the system by the following 5-rule Takagi Sugeno fuzzy model. Table 1 shows the membership function labels used in the rules.

$$\text{If } \alpha \text{ is } Z \text{ Then } \dot{x} = A_1 x + B_1 u$$
$$\text{If } \alpha \text{ is PS or NS Then } \dot{x} = A_2 x + B_2 u$$
$$\text{If } \alpha \text{ is PM or NM Then } \dot{x} = A_3 x + B_3 u$$
$$\text{If } \alpha \text{ is PL or NL Then } \dot{x} = A_4 x + B_4 u$$

$$\text{If } \alpha \text{ is PVL or NVL Then } \dot{x} = A_5 x + B_5 u$$

$$\text{where } x = \begin{bmatrix} r \\ \dot{r} \\ \alpha \\ \dot{\alpha} \end{bmatrix} \quad \dot{x} = \begin{bmatrix} \dot{r} \\ \ddot{r} \\ \dot{\alpha} \\ \ddot{\alpha} \end{bmatrix} \quad A_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-mga_i}{\frac{J}{R^2}+m} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad B_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

for $i = 1, 2, \ldots, 5$ and with $a_1 = 0$, $a_2 = 12\sin(15\pi/180)/\pi$, $a_3 = 3\pi$, $a_4 = 4/(\sqrt{2}\pi)$, $a_5 = 3\sqrt{3}/(2\pi)$

By utilizing the concept of Parallel Distributed Compensation (PDC), the following 5 rules are designed for the controller.

$$\begin{aligned} \text{If } \alpha \text{ is } Z \text{ Then } u &= k_1 x \\ \text{If } \alpha \text{ is PS or NS Then } u &= k_2 x \\ \text{If } \alpha \text{ is PM or NM Then } u &= k_3 x \\ \text{If } \alpha \text{ is PL or NL Then } u &= k_4 x \\ \text{If } \alpha \text{ is PVL or NVL Then } u &= k_5 x \end{aligned}$$

By choosing the eigen values [-2+2i, -2-2i, -0.5, -1], the controller gains are obtained as

$$K_1 = [0.5714 \quad 2.0000 \quad 14.5 \quad 5.5] \quad K_2 = [0.5780 \quad 2.0230 \quad 14.5 \quad 5.5]$$
$$K_3 = [0.5984 \quad 2.0944 \quad 14.5 \quad 5.5] \quad K_4 = [0.6347 \quad 2.2214 \quad 14.5 \quad 5.5]$$
$$\text{and} \quad K_5 = [0.6910 \quad 2.4184 \quad 14.5 \quad 5.5]$$

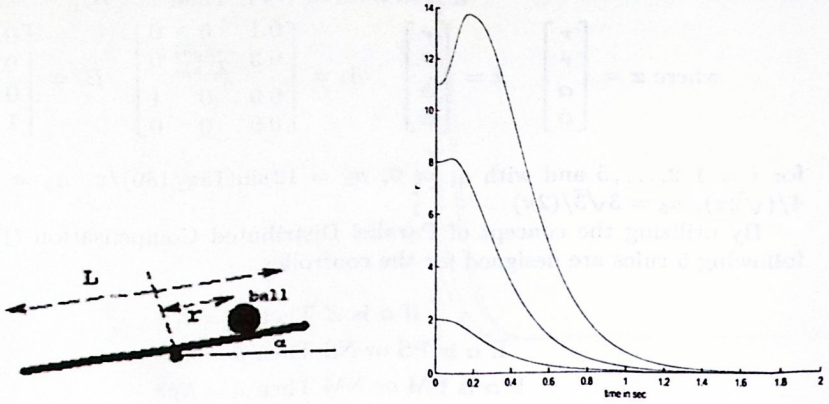
Figure 2b shows the plot of r starting from various positions. Figure 3a, 3b, 3c, and 3d shows the plot of $\dot{r}$, α, $\dot{\alpha}$ and u respectively.

Table 1. Membership function labels for α

Label	Left	Center	Right	Label	Left	Center	Right
NVL	-75	-60	-45	NL	-60	-45	-30
NM	-45	-30	-15	NS	-30	-15	0
Z	-15	0	15	PS	0	15	30
PM	15	30	45	PL	30	45	60
PVL	45	60	75				

4 Conclusion

This paper deals with the survey of Takagi-Sugeno fuzzy model and its controller. Two problems namely, inverted pendulum model and ball-and-beam model have been simulated and the results are shown. This survey and simulations using



(a) Ball and Beam system (b) Response of r

Fig. 2. Ball and Beam system and response of r

Matlab are used to teach the design of Takagi-Sugeno fuzzy controllers for undergraduate courses.

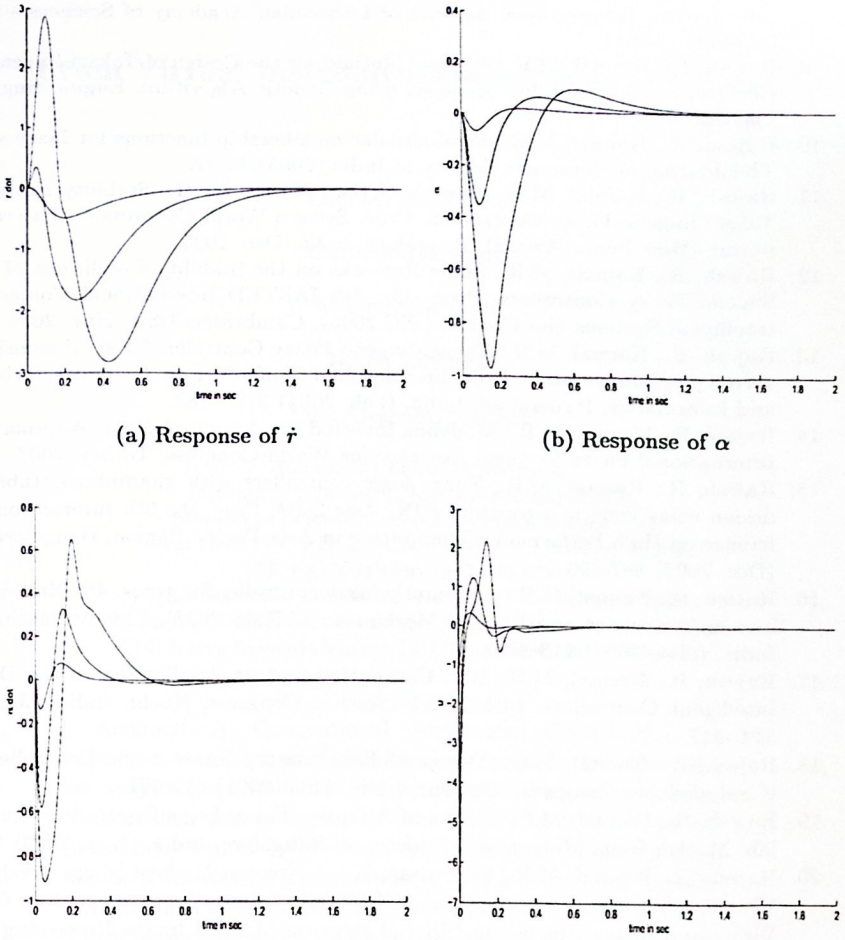
Implementations using fuzzy logic controllers can work well without having to construct any mathematical model of the process or plant. Fuzzy logic control yields results superior to those using conventional control algorithms and their applications can also lead to reduced development cost. In many cases, fuzzy control can be used to improve existing controller systems by adding an extra layer of intelligence to the current control method.

5 Acknowledgement

The authors would like to thank the reviewers for their valuable comments for improving the quality of the paper. The authors are thankful to all the staffs of the Department of Computer Science and Engineering, Bharathiar University, India.

References

1. Bowyer, K., Stockman, G., Stark, L.: Themes for improved teaching of image computation. IEEE Trans. Education 43(2) (May 2000) 221-223.
2. Jurado, F., Castro, M., Carpio, J.: Experiences with fuzzy logic and neural networks in a control course. IEEE Trans. Education 45(2) (May 2002) 161-167.
3. Kaimal, M.R., Rajesh, R., Mohanlal, P.P.: An Optimal Rule Generation Scheme using Genetic Algorithm for the Design of a Fuzzy Logic Controller. PARITANTRA, Journal of System Society of India 8(1) (Aug. 2003) 20 - 25.



(a) Response of $\dot{r}$

(b) Response of α

(c) Response of $\dot{\alpha}$

(d) Response of u

Fig. 3. Simulation results of Ball and Beam system

4. Klee, H., Dumas, J.: Theory, Simulation, Experimentation-An integrated apprach to teaching digital control systems. IEEE Trans. Education 37(1) (Feb 1994) 57-62.
5. Lam, H.K., Leung, F.H.F., Tam, P.K.S.: A linear matrix inequality approach for the control of uncertain fuzzy systems. IEEE Cont. Sys. Mag. (Aug. 2002) 20-25.
6. Matlab Version 6.0 User Manual, MathWorks, Inc. Natick, MA, 2000.
7. Rajesh, R., Kaimal, M.R.: T-S Fuzzy Model with Nonlinear Consequence and PDC Controller for a Class of Nonlinear Control Systems. Applied Soft Computing Journal - in press (avilable online).

8. Rajesh, R., Kaimal, M.R.: Variable gain Takagi-Sugeno fuzzy logic controllers. Informatica, International Journal of Lithuanian Academy of Sciences - in press (avilable online).
9. Rajesh, R., Kaimal, M.R.: A Novel Method for the Design of Takagi-Sugeno Fuzzy Controllers with Stability Analysis using Genetic Algorithm. Engineering Letters - accepted
10. Rajesh, R., Kaimal, M.R.: Semi-circular membership functions for fuzzy systems. The Journal of Computer Society of India (2003) 40-47.
11. Rajesh, R., Kaimal, M.R.: Bounds of the Solution for the Stability Equations of Takagi-Sugeno Fuzzy Controllers. Proc. Second World Congress on Lateral Computing (Best Paper Award), Bangalore, India, Dec. 2005.
12. Rajesh, R., Kaimal, M.R.: Some Remarks on the Stability Conditions of Takagi-Sugeno Fuzzy Controllers. Proc. The 8th IASTED International Conference on Intelligent Systems and Control (ISC 2005), Cambridge, USA, Nov. 2005.
13. Rajesh, R., Kaimal, M.R.: Takagi-Sugeno Fuzzy Controller for six dimensional inverted pendulum system. Proc. International Conference on Systemics, Cybernetics and Informatics, Hyderabad, India, (Feb. 2004) 278 - 283.
14. Rajesh, R., Kaimal, M.R.: Modeling Inverted Pendulum - A Novel Approach. Proc. International Fuzzy Systems Associations World Congress, Turkey, 2003.
15. Rajesh, R., Kaimal, M.R.: Fuzzy logic controllers with guaranteed stability - A design using genetic algorithm. HPC Asia 2002, Proc. the 6th International Conference on High Performance Computing in Asia-Pacific Region, Bangalore, India, (Dec. 2002) 387-390.
16. Rajesh, R., Kaimal, M.R.: Automatic fuzzy controller for gyros. 4th National Seminar on Aerospace and Related Mechanism (ARMS 2002), Thiruvananthapuram, India, (Nov. 2002) 113-118.
17. Rajesh, R., Kaimal, M.R.: Soft Computing and its Application to the Design of Intelligent Controllers. 14th Kerala Science Congress, Kochi, India, (Jan. 2002) 514 -517.
18. Rajesh, R., Kaimal, M.R.: Design of Evolutionary Fuzzy Logic Controllers. 13th Kerala Science Congress, Thrissur, India, (Jan. 2001) 668-671.
19. Rajesh, R., Kaimal, M.R.: Design of Adaptive Fuzzy Logic Controllers Using Matlab. Matlab India Millennium Conference, Bangalore, India, (Nov. 2000) 167-180.
20. Rajesh, R., Kaimal, M.R., Srinivasan, K.: A Note on Medical Image Analysis and Visualization using Matlab. Proc. ICGST and International Journal on Graphics Vision and Image Processing (Special Issue on Medical Image Processing) (2006) 352-355.
21. Takagi, T., and Sugeno, M.: Fuzzy identification of systems and its applications to modleing and control. IEEE Trans. Syst. Man and Cyber. 15 (1985) 116-132.
22. Tanaka, K. and Sugeno, M.: Stability analysis and design of fuzzy control systems. Fuzzy Sets and Systems 45(2) (1992) 135-156.
23. Wang, H.O., Tanaka, K., Griffin, M.: Parallel distributed compensation of nonlinear systems by Takagi-Sugeno Fuzzy Model. IEEE .... (1995) 531-538.
24. Wang, H.O., Tanaka, K., Griffin, M.: An approach to fuzzy control of nonlinear systems - stability and design issues. IEEE Trans. on fuzzy systems (1996) 14-23.

From Virtual Instruments to Virtual Laboratory

Luis M. Fernandez[1]

Guillermo Alvarez[1]

Miguel Machirán[1]

Ileana Sanchez[1]

Luis Alejandro Sánchez[2]

[1]Automatic Control Department
Institute of Cybernetic, Mathematic and Physic.
Minister of Science, Technology and Environment.
Street 15 #551 C y D. Zip Code 10400. C. Havana. Cuba.
(537)-8320319 Fax :537-333373
[2]Higher School of Computing, Nacional Polytechnic Institute
Av. Juan de Dios Batiz s/n casi esq. Miguel Othon de Mendizábal.
Col. Nueva Industria Vallejo CP 07738. MexicoCity, Mexico.

luis@icmf.inf.cu.

Abstract: A Conventional Automatic Control Lab is expensive because it needs industrial equipment. Space requires are very large. Installation and verification time are very long. Maintenance cost is expensive too. Virtual instrument concept (created by James J. Truchard and Jeffrey L. Kodosky from National Instrument Corporation, USA) is a low cost solution to implement virtual laboratory. This concept is using by us in a Virtual Laboratory (VL). Our VL is at Institute of Cybernetic, Mathematic and Physic. We offer training support in computer science, automation, instrumentation and digital network communication for industries and universities.

1. Introduction.

Teaching is one technology transference path because it takes the liberty offer accumulated experience to futures engineers.

Automation's techniques using computers occupies an important place in present and in immediate future of industry.

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 141-150

Majority of computer control techniques take place and passes by university laboratories before arrive in industry.

The cost of necessary devices to equip an automation laboratory is very expensive if we want to prepare students for actual industry or if we want to offer training for industry engineers. Virtual Instruments (VI) requires an important initial inversion but its can help us to reduce considerably effective cost. Let me to show a simple example.

A two channels oscilloscope, a digital multimetre and a signal generator (basic elements of any laboratory), has an approximate costs of 20 000 USD meanwhile a Personal Computer (PC), an interface card (with input-output digital-analogue signals possibility) and a LabWindows/CVI license (for VI development) can cost approximately 8 000 USD.

The difference is almost 12 000 USD. It can help us to think about laboratories equipped with VI. Laboratories equipped with VI will be denominates Virtual Labs. The virtual sense of VL isn't given only by fact that VIs are used. The fact is also that laboratory can be employed for different training subjects. If we only modify software on the PC then our automation LV can be transformed in an electronic lab or in measuring lab, or in networks lab, etc. The VL with adequate software can change from programming languages to digital network communication. This versatileness contributes to reduce laboratory effective costs and raise its flexibility of uses. Then we consider this laboratory in practice as a VL.

Another important advantage of VL is given by space saving and cheap maintenance expenditure. Software is the instrument then we are saving space because software is store in digital supports while that oscilloscopes, signal generators, multimetre, indicators, etc. requires many more real spaces and cleaning.

Cheap maintenance expenditure is given by equipment used. In this case, it is used as hardware PCs and interface cards. Its have more economic maintenance that real instruments. Then we can say that effective cost is reducing with VL.

Additionally, VI is easy updating because necessary additional code to join new functions to instruments is very little. But updating of real instrument signifies a new one.

Finally, VI can be prepared in order to process information obtained up to ours needs. It is possible to give specific solutions to specific problems. Real instruments were creating by producers to give general service. They resolve only considerate problems by producers. Many times these are not all ours problems.

Undoubtedly changes take places on concepts about what is a laboratory to teach ?, papers from [1], [2], [3], [4], [5], all presented in the San Francisco 13th IFAC Congress. [6] and [7] both presented in the Beijing 14th IFAC Congress. [8] and [9], both presented in the Barcelona 15th IFAC Congress, [10], [11] [12], all presented in the Prague 16th IFAC Congress, shown interest level on this area.

In this paper, I have dedicated first part to present the methods and materials employed in different experiments of a VL; second part is to show one of these experiments, principally: developed VI and obtained results. Next part contains conclusions and the last one contains references.

2 Methods and Tools

Laboratory equipment was carefully selected. It was settling with 6 work places with capacity of until 2 students per place.

In each case we employ the following tools:

- PC Pentium III 766 MHz, 128 Mbytes RAM, HD 40 Gbytes, SVGA.
- Microsoft Corporation Windows 2000 Professional.
- SCADA programming in C.
- Microcontroller digital communication between process and PC, RS-232C.

VL project was supported by Nuclear Energy Agency of Cuba.

Many different experiments were designing for this VL [13], [14], [15] y [16]; we present here each one briefly.

2.1. Fast process, DC motor.

This experiment consists in a DC motor with tachometer coupled and communication board, based in PIC microcontroller, as show in figure 1.

SCADA is placed in PC and it solves human – machine communication problem, see figure 2.

Target is to use a PID digital control of fast process. Students can take measurements or dc motor parameters, make identification and design control with digital PID.

2.2 Slow process, Low Temperature furnace.

This experiment consists in little furnace, see figure 3 and communication board, based in PIC microcontroller.

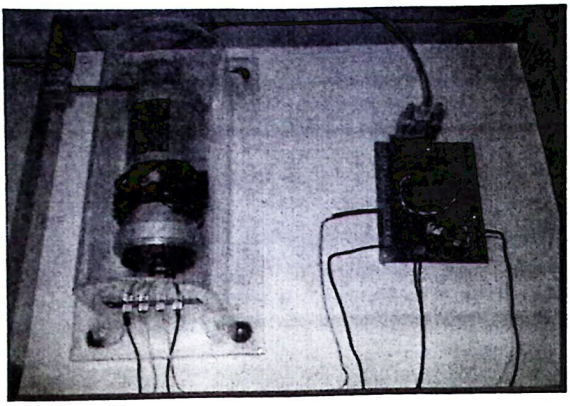


Fig. 1. DC motor and communication card.

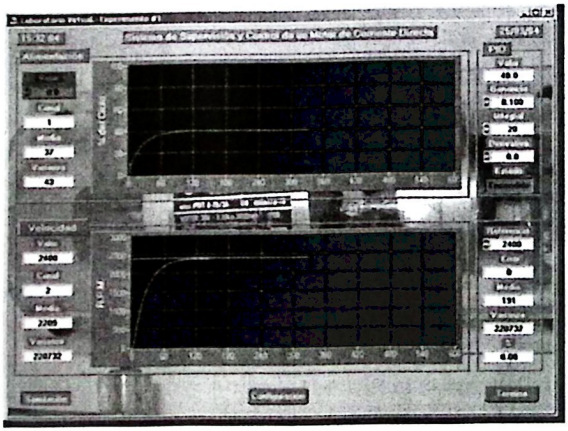


Fig. 2. DC motor SCADA

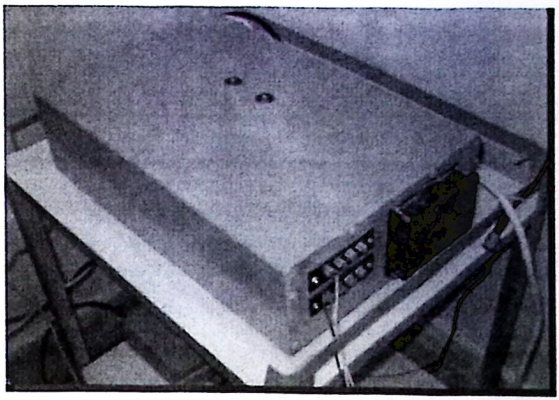


Fig. 3. Furnace

SCADA is placed in PC and it solves human – machine communication problem, see figure 4.

Target is to use a PID digital control of slow process. Students can take measurements or furnace parameters, make identification and design control with digital PID.

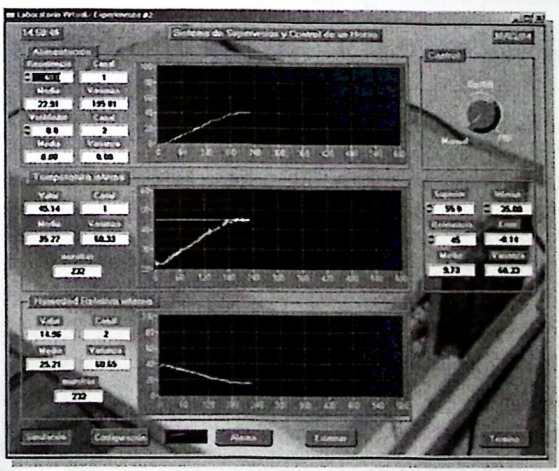


Fig. 4. Furnace SCADA.

2.3. Sequential control, Traffic Control network.

Experiment consists in a Traffic Control network locates in Salvador Allende and Infanta interception in Havana City, see figure 5. Communication with PC is across a board, based in PIC microcontroller.

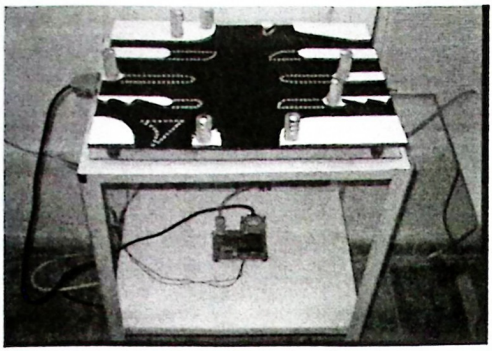


Fig. 5. Network for Traffic Control.

SCADA is placed in PC and it solves human – machine communication problem, see figure 6.

Target is to find a solution for each traffic problem in the net (priorities, collisions, wait time, etc.).

2.4. Advanced technology and data post processing, Vibration studies.

This experiment consists in 3 step motors; see figure 3, one optical fibre sensor and one communication board, based in PIC microcontroller.

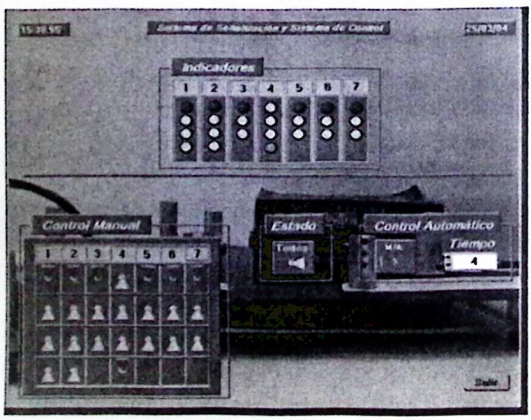


Fig. 6. SCADA for Traffic Control Network

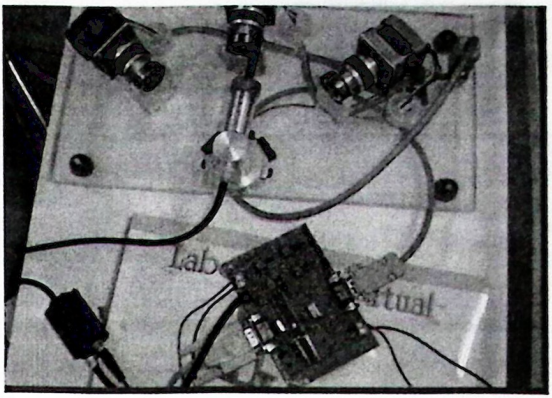


Fig. 7. Vibration studies.

SCADA is placed in PC and it solves human – machine communication problem, see figure 8.

Target is to study of cutting edge of technology with optical fibre sensor and mathematical tools for data post processing in bearing vibration problem.

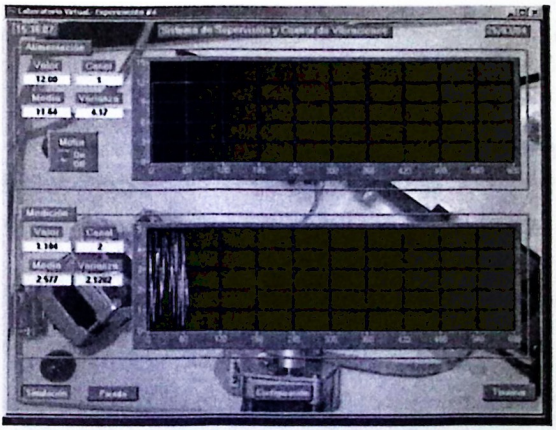


Fig. 8. SCADA for vibration studies.

2.5 Complex process, Vacuum control.

This experiment consists in one AC motor, one vacuum pump, one cylindrical tank, one valve with step motor for positioning, see figure 9.

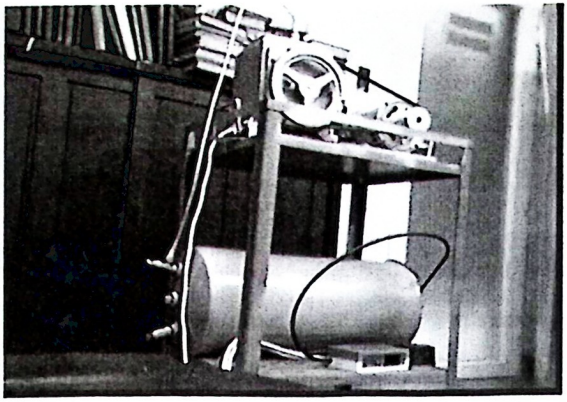


Fig. 9. Complex process.

SCADA is placed in PC and it solves human – machine communication problem, see figure 10.

Target is to use a PID digital control for complex process. Students can take measurements, make identification and design control with digital PID.

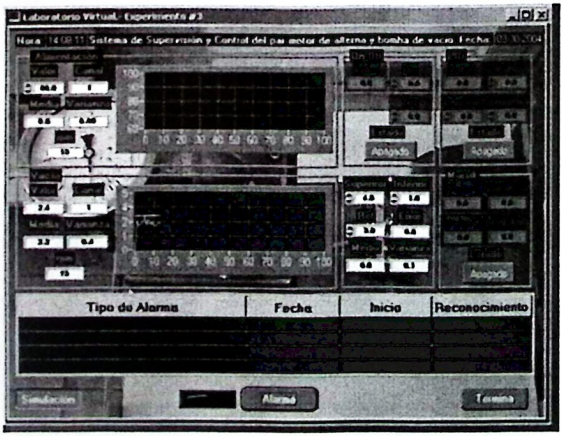


Fig. 10. SCADA for complex process.

3. Results.

Each experiment has prepared students to develop different abilities in measurements and control techniques.

VL can be used for training in hardware design, industrial digital network protocols, software develop, data acquisition techniques, post processing data techniques, identification and control.

Training is doing with middle level students, university students, postgraduate students. For Curricula of Masters and Doctoral courses, experimental activities were made.

4. Conclusions

Virtual Laboratory is a low cost solution in relation with traditional laboratory. This economic solution can be employed by non rich institutions.

A Virtual Laboratory with capacity until 12 students by group has been created at *Institute of Cybernetic, Mathematic and Physic, Havana, Cuba.* Then it is possible to use this solution for university laboratory.

Many different Virtual Instruments has been developed to experimentation in the laboratory conditions. This instrumentation type opens the possibility of Virtual Instrument interchange between institutions. Actually, this experience exists.

Automatic control with Virtual Instrument has been employed in many different country industries. This path reduce gap between university and industry and smooth new engineer adaptations.

Therefore, there is a solution to bring near industry problems at university. Moreover this can help to create Virtual Instruments at university for industry.

References

1. Bohus, C., L. A. Crowl, B. Aktan and M. H. Short. (1996). Running control engineering experiments over the Internet. 13th Triennial World Congress, San Francisco. USA.

2. Derighetti, M., A. H. Glattfelder, G. Peretti, X. Qiu, L. Rickenbach, W. Schaufelberger and J. Wang (1996). Laboratory experiments for integrated basic control course. 13th Triennial World Congress, San Francisco. USA.

3. Kozura, J., H. Kaufman, J. Raber, R. Shamayer, B. W. Bequette and D. O'Loughiin. (1996). Computer Animated Control Laboratory Emulations Modules. 13th Triennial World Congress, San Francisco. USA.

4. Schumann, R., M. Kluver, D. Matko and D. Vitkute. (1996). A New Concept for computer aided learning in the control laboratory. 13th Triennial World Congress, San Francisco. USA.

5. Vagners, J. (1996). Integrated Laboratory experience for control education. 13th Triennial World Congress, San Francisco. USA.

6. Bark A, Egardt B, Kristiansson B (1999) A Web-Based Course in Automatic Control for Engineers in Industry, 14th Triennial IFAC World Congress, Beijing. China.

7. Horá oek Petr (1999) LABORATORY EXPERIMENTS FOR CONTROL THEORY COURSES: A SURVEY, 14th Triennial IFAC World Congress, Beijing. China.

8. Jukka L, Mikko H, Tero K (2002). A Distance Learning Course in Modelling and Simulation, 15th Triennial IFAC World Congress, Barcelona. Spain.

9. Arnanz R, Miguel Angel Pacheco, Antonio Mendoza, José Bernardez, José R. Perán González (2002) Development of a Distant AC Motor Laboratory, 15th Triennial IFAC World Congress, Barcelona. Spain.

10. Martin C, Urquia A, Dormido S (2005). Object-oriented Modeling of Virtual Laboratories for Control Education, 16th Triennial IFAC World Congress, Prague. Czech Republic.

11. Feng Y, Rong G (2005). Virtual Plant Laboratory System of Process Industries for Education, 16th Triennial IFAC World Congress, Prague. Czech Republic.

12. Petrosyan E, Vardan M. (2005). A Study on Control Problems in Online Distance Education, 16th Triennial IFAC World Congress, Prague. Czech Republic.

13. Fernández, L. M. and C. Gutiérrez. (1997). Instrumentación Virtual para la Enseñanza. Conferencia Internacional CIMAF'97. La Habana. Cuba.

14. Fernandez L. M. (1998). Virtual Instruments, a low cost solution to Education Laboratory. V Symposium on Low Cost Automation, Shenyang, China.

15. Fernández L. M. et al (2004). Laboratorio Virtual para el entrenamiento del personal de la industria. XI Congreso Latinoamericano de Control Automático. La Habana, Cuba.

16. Gutiérrez C. et al (1997). La Instrumentación Virtual en el mejoramiento de la enseñanza de cursos de mediciones e ingeniería de control : diseño experimental, prototipo y resultados en México". Instrumentaion Newsletter. Vol 9 Number 3 Autumn 1997. Austin, Texas, USA.

Environmental Noise Monitoring. 1st. Stage: Aircrafts Noise Patterns Recognition

Luis P. Sánchez[1], Lucrecia Pérez Echazabal,[2] and Luis Alejandro Sánchez Pérez[3]

[1]Center for Computing Research, National Polytechnic Institute

[2] Instituto Tecnologico y de Estudios Superiores de Monterrey, Mexico City

[3]Higher School of Computing, National Polytechnic Institute

Av. Juan de Dios Batiz s/n casi esq. Miguel Othon de Mendizabal, Col. Nueva Industrial Vallejo. CP 07738. Mexico City, Mexico

lsanchez@cic.ipn.mx

Abstract. The noise level in a city has fluctuations between 50 dB (A) and 100 dB (A). It depends on the population density and its activity, commerce and services in the public thoroughfare, terrestrial and aerial urban traffic, of the typical activities of labor facilities and used machinery, which give varied conditions that must be faced of diverse ways within the corresponding normalization. The sounds or noises that exceed the permissible limits, whichever the activities or causes that originate them, are considered events susceptible to degrade the environment and the health. This paper is a task within an environmental noise monitoring system and it presents the analysis of the aircraft noise signals and a method to identify them. The method uses processed spectral patterns and a neuronal network feed-forward, programmed by means of virtual instruments.

1 Introduction

Much of this work involves the collection and analysis of large amounts of aircraft noise data from the Noise and Track Keeping systems (NTK) installed at airports. Like any other measured quantity, aircraft noise measurements are subject to some uncertainty, which can influence the quality of the final measured result [1].

The uncertainty contributions for a typical noise study can be considered in two groups. The first group includes the components of uncertainty associated with the measurement of aircraft noise at a particular monitoring location. The second group includes the components of uncertainty associated with any subsequent data analysis that may be carried out [1], [2], [3]. The overall accuracy of any type of measurement is limited by various sources of error or uncertainty. Components of uncertainty can essentially be classified as either *random* or *systematic* in nature. When making a series of repeated measurements, the effect of the former is to produce randomly different results each time, which are all spread or scattered around an average (mean) value. In contrast, systematic components of uncertainty cause the measurement to be consistently above or below the true value. For example, when measuring the time with a watch that has been set 1 minute slow, there will be a systematic error (or bias)

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 151-160

in all the measurements. In a well-designed measurement study, the systematic components of uncertainty should generally be smaller than the random components [1], [4]. Possible sources of uncertainty for aircraft noise measurements include not only the noise instrumentation itself, but also variations in the noise source and propagation path, meteorological variations, the local environment at the measurement site, and also any variance due to data sampling - all of these individual uncertainty components can influence the quality of the final measured result [1]. An internationally accepted procedure for combining and expressing measurement uncertainties is given in the ISO Guide to the Expression of Uncertainty in Measurement [5], [6], [7].

The used aircraft noises in this work have been acquired by means of archives of sounds, whose measurements were made fulfilling the norms mentioned previously, with frequencies of sampling of 22050 Hertz, monophonic, during 24 seconds. In general, this interval, is greater than to aircraft takeoff time, or greater to the time in which the produced noise affects the zones near an airport.

For a same aircraft, several archives of sounds were used, taken for different meteorological conditions and microphone orientation. Two types of microphones were used, which can be considered typical for these measurements. In addition, the sounds were reproduced with three different sound cards.

When the environmental noise monitoring system is in operation, norms of measurements and calibration will be elaborated

2 Diagrams and description of typical architecture of an aircraft noise monitoring station [8], [9].

Generally, a noise monitoring complex system detects, identifies and analyses the noise produced by arriving and departing aircrafts. The Fig. 1 presents a typical architecture of aircraft noise monitoring stations

The noise monitoring system (NMS) measures aircraft noise according to defined criteria. The first step of the system is the collection of the detected aircraft noise, the second the attribution of the noise to a specific aircraft movement. To perform the correlation of the aircraft noise, additional information is necessary, which will be described later.

2.1 EMU - Environmental Monitoring Unit

The EMU consists of

- a digital microphone for noise measurement
- a local unit for data backup
- a modem for transmitting data to the central processing system.

2.1.1 Microphone unit

Each unit is mounted at the end of a mast and equipped with a digital microphone, an anti-wind and bird guard and a lightning arrestor. The microphone captures the

analogue noise signal and performs the critical conversion of the signal immediately at the microphone head, and transmits the noise data in digital form to the EMU's electronics. The immediate conversion to a digital signal provides a higher immunity to interference. The unit guarantees an omni directional detection of noise with high reception qualities. The 5 local units are synchronized by the central system GPS clock.

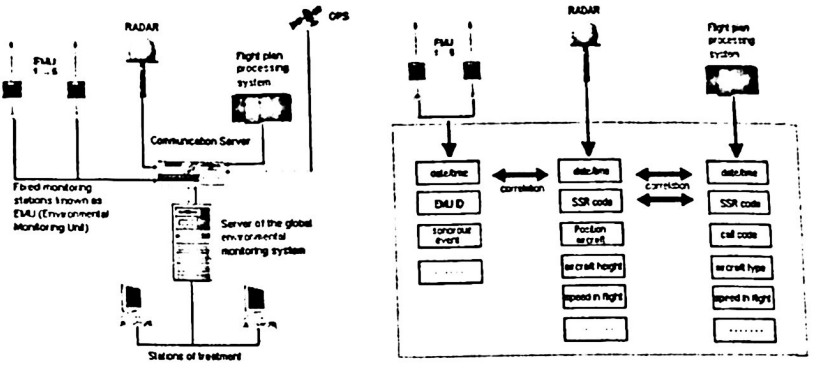


Fig. 1 Typical architecture of aircraft noise monitoring stations and correlation principle

2.2 Central processing system

All the data collected via network or modem from the airport radar, the flight plan processing system and the EMUs are put through to the central processing system which consists of **Communication Server** and **Global Environment Monitoring System**. The Communication Server collects:

- the noise events of the five EMUs,
- the radar aircraft tracks,
- the flight plans from FDP (Flight Data Processor),
- the GPS (Global Positioning System) to guarantee the synchronization of the noise monitoring system.

The Global Environment Monitoring System processes and correlates the data acquired by the Communication Server in order to identify an aircraft that produced a noise event. The essential processing in the Global Environment Monitoring System consists of:

- correlating the real time data from the Communication Server server;
- archiving the acquired data, and generating reports about individual events, daily, monthly or annual summaries.

Identification of the noise event: The EMU continuously analyses the incoming noise signal to identify the source of noise. By using various detection algorithms it is possible to identify noise generated by an aircraft flying past, known as event. The process of identifying a noise event is based on threshold and time change criteria. The incoming data are noise events, aircraft flight plans and Radar information. The

correlation principle is observed in the right side of Fig 1.

3. Aircrafts noise patterns

The goal of this development stage of the environmental noise monitoring system is to make aircrafts noise signals analysis, that allows to create a method of aircraft noise patterns recognition, that will do possible to identify types of aircrafts. These types of aircrafts can be of helix, turbojet and reaction. On the other hand, it is possible to classify the aircrafts as long reach (high power), medium reach (medium power) and short reach (low power). Committees of Aerial transport and environmental propose an aircraft classification based on the level of noise emission.

The proposed common classification of aircraft is based on the principle that the aircraft operator should pay a fair price that should be proportional to its noise impact, independently of the weight of the aircraft or of the transport service rendered. Such data would make it possible to recognize the environmental merits of larger aircraft, even if these aircraft are noisier in absolute terms when compared to smaller aircraft. Therefore the proposal contains a discretionary provision for additional information to be given to the public concerning the noise productivity of heavier aircraft. This is to ensure that the concept of noise productivity is well understood [10]. The presented method recognizes specific aircraft type.

3.1 Aircraft noise signals analysis

As example, we will present noise signals of some aircrafts. The used aircraft noises in this work have been acquired by means of archives of sounds, whose measurements were made fulfilling the norms mentioned previously, with frequencies of sampling of 22050 Hertz initially, and later to 11025 Hz, monophonic, during 24 seconds. En general, this interval, is greater than to aircraft takeoff time, or greater to the time in which the produced noise affects the zones near an airport (Fig. 2 and 3).

For all used aircraft noises the typical form of the amplitude spectrum is observed from 0 to 5000 Hertz, for this reason, we used a sampling frequency of 11025 Hz, in order to reduce the number of taken samples in 24 seconds (264600 samples). The amplitude spectrum has 132300 harmonics, with Δf = 0.04167 Hz.

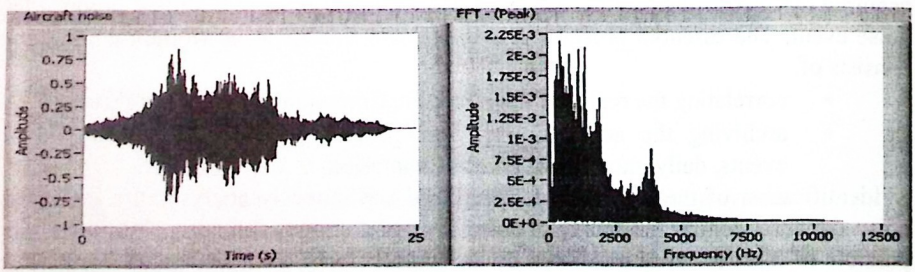


Fig. 2 Noise of Falcon aircraft taking off, with sampling frequency of 22050 Hz.

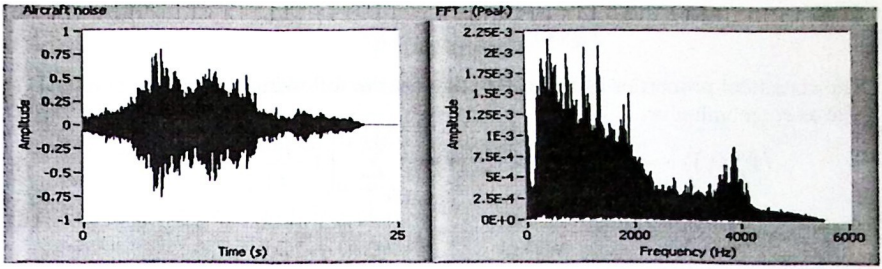


Fig. 3 Noise of Falcon aircraft taking off, with sampling frequency of 11025 Hz.

3.1.1 Reduction of the spectral resolution

It is necessary to reduce the spectral resolution because of the following reasons:

1. The amplitude spectrum has 132300 harmonics and its processing will be very complex.
2. It is only of interest the spectral form.

The following hypotheses are presented:

1. Any reduction method of spectral resolution introduces a tolerance in the initial and final times within the measurement interval of aircraft noise.
   For example, a feedforward neural network is trained with one noise pattern which was acquired from zero seconds from the aircraft takeoff until 24 seconds later. In run time, the aircraft takeoff noise is acquired from 5 seconds until 24 seconds. This time displacement of 5 seconds affects little the spectral form if its spectral resolution has been reduced
2. A median filter (moving average filter) creates a typical form of the aircrafts takeoff noises spectrums.
3. The decimation of average spectrum, with a rate X, conserves the spectral form of aircrafts takeoff noises.

3.1.2 Spectral estimation

In the present work is used the Bartlett-Welch method [11] for spectral estimation. The Bartlett method consists on dividing the received data sequence into a number K, of non-overlapping segments and averaging the calculated Fast Fourier Transform.

It consists of three steps [12]:

1. The sequence of N points is subdivided in K non overlapping segments, where each segment has length M.

$$x_i(n) = x_i(n + iM),\ i = 0,1,...,K-1,\ n = 0,1,...,M-1 \tag{1}$$

2. For each segment, periodogram is calculated

$$\hat{P}_{xx}(f) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_i(n) e^{-j2\pi fn}\right|^2,\ i = 0,1,...,K-1 \tag{2}$$

3. The periodograms are averaged for the K segments and the estimation of the Bartlett spectral power can be obtained (in this work, we don't use the spectral power) as:

$$\hat{P}_{xx}^{B}(f)=\frac{1}{K}\left|\sum_{i=0}^{K-1}\hat{P}_{xx}^{(i)}(f)\right|^{2} \tag{3}$$

The statistical properties of this estimation are the following ones:

The average value is:

$$\left\langle \hat{P}_{xx}^{B}(f)\right\rangle=\frac{1}{K}\sum_{i=0}^{K-1}\overline{\hat{P}_{xx}^{(i)}(f)}=\overline{\hat{P}_{xx}^{(i)}(f)}=\sum_{m=-(M-1)}^{M-1}\left(1-\frac{|m|}{M}\right)r_{xx}(m)e^{-j2\pi fm}$$

$$=\frac{1}{M}\int_{-1/2}^{1/2}P_{xx}(a)\left(\frac{\operatorname{sen}\pi(f-\alpha)M}{\operatorname{sen}\pi(f-\alpha)}\right)^{2}d\alpha, \tag{4}$$

donde $\frac{1}{M}\frac{\operatorname{sen}\pi(f-\alpha)}{\operatorname{sen}\pi(f-\alpha)}=W_{B}(f)$ it is the frequency characteristic of the Bartlett window: $w_{B}(m)=\begin{cases}\left(1-\frac{|m|}{N}\right), & |m|\le M-1\\ 0, & |m|>M-1\end{cases}$ (5)

The true spectrum is convolutioned with the frequency characteristic of the Bartlett window $w_{B}(m)$. Reducing the longitude of the data window of N points to M=N/K, it results in a window whose spectral wide has been increased by the factor k. Consequently, ***the frequency resolution has decreased for the factor k,*** in exchange for a variance reduction.

The variance of the Bartlett estimation is:

$$\operatorname{var}\left[\hat{P}_{xx}^{B}(f)\right]=\frac{1}{K^{2}}\sum_{i=0}^{K-1}\operatorname{var}\left[\hat{P}_{xx}^{(i)}(f)\right]=\frac{1}{K}\operatorname{var}\left[\hat{P}_{xx}^{(i)}(f)\right]==\frac{1}{K}P_{xx}^{2}(f)\left[1+\left(\frac{\operatorname{sen}2\pi fM}{M\operatorname{sen}2\pi f}\right)^{2}\right] \tag{6}$$

Welch Method [11], [13], [14]: unlike in the Bartlett method, the different data segments are allowed to overlap and each data segment is windowed.

$$x_{i}(n)=x(n+iD),\ n=0,1,...,M-1,\ i=0,1,...,L-1 \tag{6}$$

Where iD is the point of beginning of the sequence i-ésima. If D=M, the segments are not overlapped. If D=M/2, the successive segments have 50% of overlapping and the obtained data segments are L=2K.

Another modification proposed by Welch to the Bartlet method consists on using a window for the data segments before calculating the periodogram. The result is the "modified" periodogram:

$$\tilde{P}_{xx}^{(i)}(f)=\frac{1}{MU}\left|\sum_{n=0}^{M-1}x_{i}(n)w(n)e^{-j2\pi fn}\right|^{2},\ i=0,1,...,L-1 \tag{7}$$

Where U is a normalization factor for power of the function window and it is selected as:

$$U=\frac{1}{M}\sum_{n=0}^{M-1}w(n) \tag{8}$$

The Welch estimation of spectral power is the average of these modified periodograms:

$$P_{xx}^{W}(f)=\frac{1}{L}\sum_{i=0}^{L-1}\tilde{P}_{xx}^{(i)}(f) \tag{9}$$

The average of the Welch estimation is:

$$\left\langle P_{xx}^{W}(f)\right\rangle=\frac{1}{L}\sum_{i=0}^{L-1}\overline{\tilde{P}_{xx}^{(i)}(f)}=\overline{\tilde{P}_{xx}^{(i)}(f)}=\frac{1}{MU}\int_{-1/2}^{1/2}P_{xx}(a)\,W(f-\alpha)\,d\alpha \tag{10}$$

$$\text{Where } W(f)=\frac{1}{MU}\left(\frac{\operatorname{sen}\pi(f-\alpha)M}{\operatorname{sen}\pi(f-\alpha)}\right)^{2}=\frac{1}{MU}\Im\{w(n)\} \tag{11}$$

The normalization factor assures that: $\int_{-1/2}^{1/2}W(f)\,df=1$

The variance of the Welch estimation is:

$$\operatorname{var}\left[\hat{P}_{xx}^{W}(f)\right]=\frac{1}{L^{2}}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\left\{\left[\overline{\tilde{P}_{xx}^{(i)}(f)\tilde{P}_{xx}^{(j)}(f)}\right]^{2}-\left[\left\langle\hat{P}_{xx}^{W}(f)\right\rangle\right]^{2}\right\} \tag{12}$$

Why Welch method is introduced?

– Overlapping allows more periodograms to be added, in hope of reduced variance.

– Windowing allows control between resolution and leakage.

The Welch method is hard to analyze, but empirical results show that it can offer lower variance than the Bartlett method, but the difference is not dramatic.

• Suggestion is that 50 % overlapping is used.

In this paper, the data segment of 264600 samples, acquired in 24 seconds, is divided in 24 segments: K=24, with 50% of overlapping, therefore, L=2K=48 overlapped data segments, later is applied the FFT (periodogram) to each segment and they are averaged.

3.2 Examples of aircrafts noise patterns

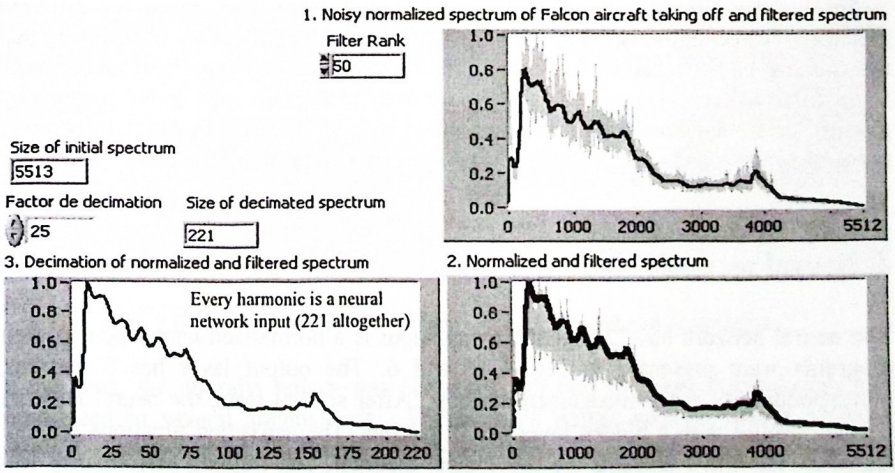


Fig. 4 Example of noise pattern of Falcon aircraft taking off (turbojet)

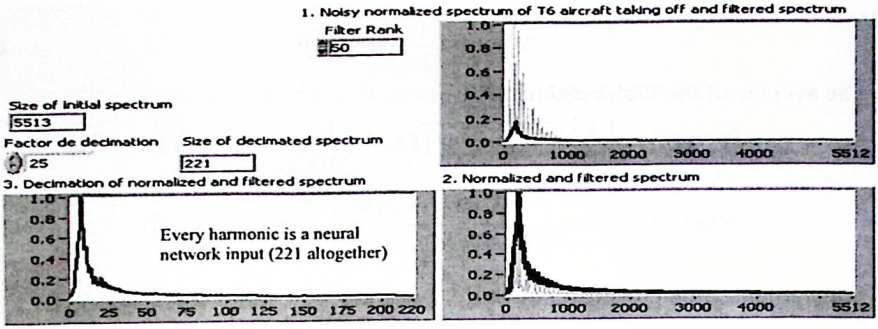


Fig. 5 Example of noise pattern of T6 aircraft taking off (helix)

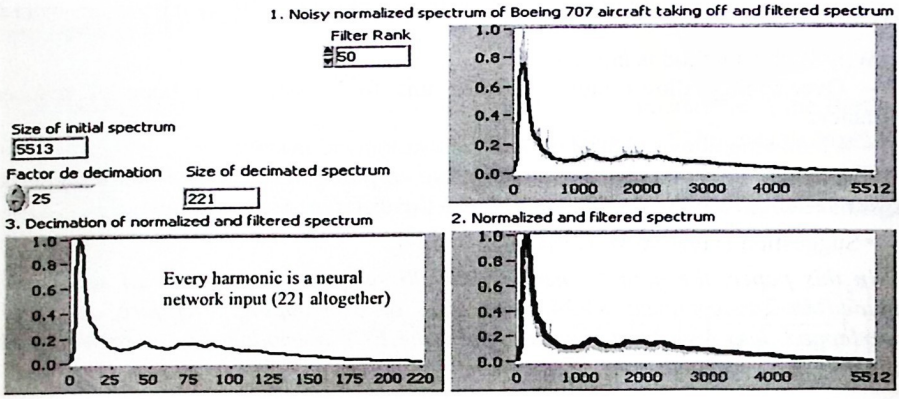


Fig. 6 Example of noise pattern of f Boeing 707 aircraft taking off (helix)

For a same aircraft, several archives of sounds were used, taken for different meteorological conditions, microphone orientation and amplification, in such way that the training patterns had distortions. In addition, the sounds were reproduced with three different sound cards. In this stage, we used 16 aircrafts type with 6 patterns by aircraft. In all, the neural network was trained with 96 patterns. In future tests, more types of aircrafts and greater amount of patterns by aircraft will be added

4. Neural network

The neural network has 221 inputs. Every input is a normalized harmonic and their diagrams were presented in Fig. 4, 5 and 6. The output layer has 8 neurons, corresponding to the 8 recognized aircrafts. After several tests, the neural network was successful with a hidden layer of 10 neurons. The activation functions are tan-sigmoid. The Fig 7 presents the topological diagram. The training performance was successful and it is presented in Fig. 8.

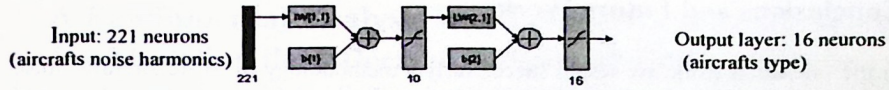


Fig 7. Neural network topological

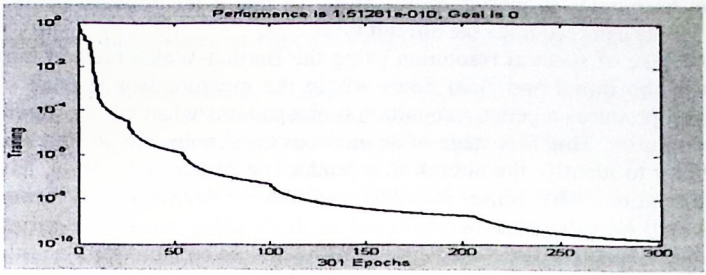


Fig 8. Training performance of neural network

4. Analysis of results

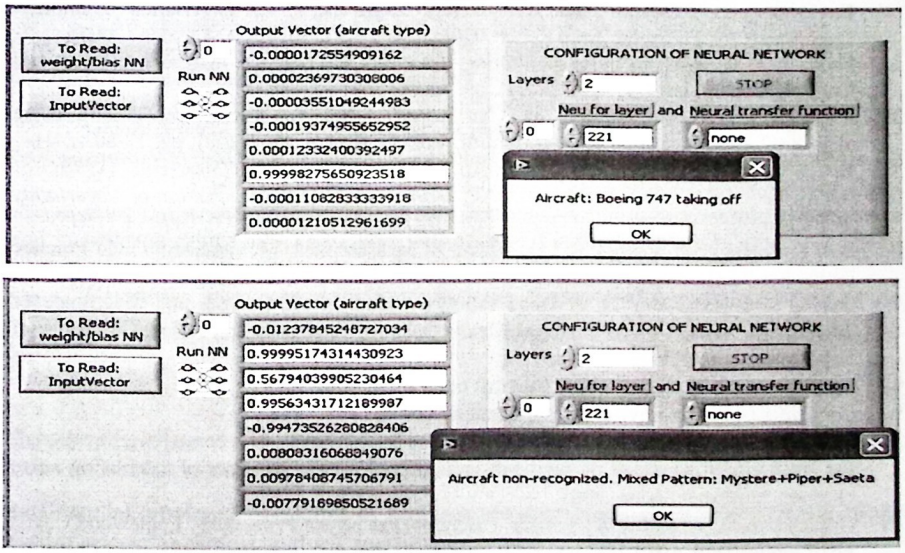


Fig. 9 Example of two tests of patterns recognition and non-recognition

In the tests, the aircrafts noises was acquired with two different microphones and reproduced in several sound cards. For aircrafts noises used in neural network training, the recognition was successful. For a aircraft non-used in the training, the neural network non- recognized a specific aircraft and presented a message with three aircrafts whose mixed noise patterns have similarity to the acquired noise (see Fig. 9).

Conclusions and Future Work

In the presented work, we tested successfully a methodology to create aircrafts noise patterns. It combine the decrease of spectral resolution, a moving average filter and decimation of average spectrum, This method allows reducing the number of significant harmonics in amplitude spectrum, so that a feedforward neural network with 221 inputs can recognize the aircraft type.

The decrease of spectral resolution using the Bartlett-Welch method introduces a tolerance in the initial and final times within the measurement interval of aircraft noise, which produces a better recognition of the patterns when the measurements can have uncertainties. This first stage of an environmental noise monitoring system tests the feasibility to identify the aircraft that produces a certain noise level, having only noise information. The noise intensity and others environmental contamination indicators will be calculated by statistical methods using noise time series. A next goal will be construct a distributed system with wireless communication with Doppler effect compensation.

References

1. White, S.: *Precision of Aircraft Noise Measurements at the London Airports.* Environmental Research and Consultancy Department. Civil Aviation Authority. Department for Transport. Ercd Report 0506. ISBN 0-117905-19-4. London (2005).
2. Holding, J. M.: Aircraft noise monitoring: principles and practice, IMC measurement and Control, Vol. 34, issue 3, April (2001), pp. 72-76.
3. Holding, J. M. and Sheldon, M.: "an expert system for aircraft recognition using acoastics", IMC measurement and Control, Vol. 34, issue 3, April 2001, pp. 77-80.
4. Doebelin, E.O.: Measurement Systems Application and Design. McGraw-Hill. (1998)
5. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization (ISO), Geneva, (1995)
6. Bell S, A Beginner's Guide to Uncertainty of Measurement, Measurement Good Practice Guide No. 11 (Issue 2), National Physical Laboratory, March (2001)
7. Craven N J and Kerry G, A Good Practice Guide on the Sources and Magnitude of Uncertainty Arising in the Practical Measurement of Environmental Noise, University of Salford, October (2001).
8. Luxembourg Airport Authority.: Aircraft noise monitoring stations. (2005)
9. Lochard Expanding Environmental Capacity (2006)
10. Kendall, M. EU proposal for a directive on the establishment of a community framework for noise classification of on civil subsonic aircraft for the purposes of calculating noise charges, European Union, 2003
11. Oppenheim, A.V., and R.W. Schafer. *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989. Pgs. 311-312.
12. Pogrebnyak, O.: Notes of classes, Center for Computing Research, National Polytechnic Institute, Mexico, 2006.
13. Welch, P.D. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms." *IEEE Trans. Audio Electroacoust.* Vol. AU-15 (June 1967). Pgs. 70-73.
14. Thompson, S.C.: Spectral Estimation of Digital Signaling Using The Welch Method. Center for Wireless Communications, Department of Electrical and Computer Engineering, University of California at San Diego (2004)

A Fundamental Analysis based Neuro-fuzzy Approach for Modeling the Effects of Crude Oil Ending Stocks Variations on Oil Price

Ali Ghaffari[1], Mansour Nikkhah Bahrami[2], Seyed Mahdiar Ghaffari[2]

[1] Mechanical Engineering Department, KN Toosi University of Technology,
322 West Mirdamad Ave., Tehran, Iran
[2] Engineering Department, University of Tehran,
North Kargar Ave., Tehran, Iran
ghaffari@kntu.ac.ir, mbahrami@ut.ac.ir, mahdiar@email.com

Abstract. Intelligent computing tools such as artificial neural network (ANN) and fuzzy logic approaches are proven to be efficient when applied individually to a variety of problems. Recently there has been a growing interest in combining both these approaches, and as a result, neuro-fuzzy computing techniques have evolved. In the last decade, various neuro-fuzzy systems have been developed. Among them, adaptive neuro-fuzzy inference system (ANFIS) provides a systematic and directed approach for model building and gives the best possible design parameters in minimum time. They have got wide acceptance for modeling many financial systems, among technical analysts. The technical analysis is only and only based on the history of the price movement. One such problem frequently encountered is the crude oil price (COP) modeling. Supply and demand are the essential factors which affect the crude oil market, such as any other financial system. One of the most important factors influencing demand is crude oil ending stocks (COES). In this paper, contrary to the classic neuro-fuzzy modeling methods, next to the time-series of the oil price, we have also considered the fundamental analysis by studying the effects of weekly COES variations as a fundamental factor. Final results show that employing COES cause better performance and less error in modeling.

Keywords: fuzzy and neuro-fuzzy systems, fundamental analysis, financial systems, crude oil ending stocks, oil price.

1 Introduction

In financial problems, two kinds of analyses are commonly used for prediction of the future prices: Technical analysis and fundamental analysis.

Technical analysis is the examination of past price movements to forecast future price direction. Technical analysts are sometimes referred to as "chartists" because they rely almost exclusively on charts for their analysis. Technical analysis is applicable to stocks, indices, commodities, futures or any tradable instrument where the price is influenced by the forces of supply and demand. Price refers to any combination of the open, high, low or close for a given commodity/security over a specific timeframe. The time frame can be based on intraday (tick, 5-minute, 15-

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 161-170

minute or hourly), daily, weekly or monthly price data and last a few hours or many years. In addition, some technical analysts include volume and/or open interest figures with their study of price action. It's been said, "A technical analyst knows the price of everything, but the value of nothing". Technicians, as technical analysts as they are called, are only concerned with two things: What is the current price? What is the history of the price movement?

The price is the end result of the battle between the forces of supply and demand for any particular item. The objective of analysis is to forecast the direction of the future price. By focusing on price and only price, technical analysis represents a direct approach. Fundamentalists are concerned with "why" the price is what it is. For technicians, the "why" portion of the equation is too broad and many times the fundamental reasons given are highly suspect. Technicians believe it is best to concentrate on "what" and never mind why. Why did the price go up? It is simple, more buyers (demand) than sellers (supply). After all, the value of any item is only what someone is willing to pay for it. Who needs to know why? You may never know why.

Aside from technical analysis, another primary approach to analyze currency market fluctuations is called fundamental analysis. Fundamental analysis is the examination of economic indicators, asset markets and political considerations when evaluating a nation's currency in terms of another. The key to fundamental analysis is to gather and interpret this information and act before the information is incorporated into the currency price. The lag time between an event and its resulting market response presents a trading opportunity for the fundamentalist [1].

Fuzzy-rule based approach in modeling is a soft computing technique which has recently received attention in financial systems. First introduced by Zadeh [15], fuzzy logic and fuzzy set theory are employed to describe human thinking and reasoning in a mathematical framework. Fuzzy-rule based modeling is a qualitative modeling scheme where the system behavior is described using a natural language [11]. The last decade has witnessed many applications of a fuzzy logic approach in financial systems forecasting [3,8,13]. The principal of designing these systems is neuro-fuzzy computing. Neural networks identify the models and adapt themselves to be able to consider the environmental variations, and fuzzy inference systems combine the human knowledge, concluding, and decision-making. The coalition of these two perfect approaches, next to the modern optimization techniques creates a new structure called neuro-fuzzy and soft computing and finally caused forming adaptive neuro-fuzzy inference systems (ANFIS).

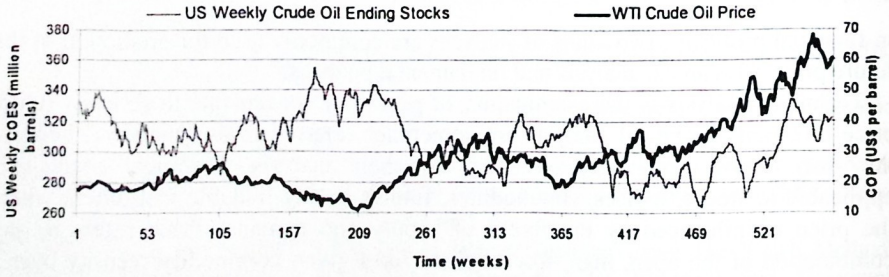


Fig. 1. The trends in WTI crude oil price and US weekly crude oil ending stocks (1995-2005).

Crude oil reserves of industrial countries such as US and China have key role in estimating the crude oil demand by fundamental analysts. In Fig. 1, we can visualize the trends in the West Texas Intermediate (WTI) crude oil price and US weekly COES from 1995 to 2005. While there is a continuous trend in the price, the relationship depicted in Fig. 1 indicates that in short- term variations both the trends parallel each other in an opposite direction. Often when US COES goes down, the price rises.

The main purpose of this paper is to try and investigate the determinants of weekly COES variations in the crude oil price on the basis of short-term horizons. For this purpose a predicting model with oil price time series inputs employed. One additional input that presents weekly COES variations added to the model, and using the error criterions we analyze the effects of this change.

2 Neuro-fuzzy Computing

Fuzzy set theory extends the degree of membership of an arbitrary element of the universe in a set to the interval [0, 1]. In the case of classical set theory there is just 0 (non-membership) and 1 (membership). If, for instance, 170 cm is the value to distinguish between tall male and short male persons in central Europe by classical sets, a height of 171 cm is assumed as tall whereas a height of 169 cm is assumed as short [7]. Fuzzy set theory permits a soft, gradual transition between the two sets, which is closer to common sense [14,16].

Hence, fuzzy sets are defined by membership functions which represent the gradual membership of all elements of the universe in the respective fuzzy set (Ross 1995). Contrary to classical sets, elements of a fuzzy set can also be members of other fuzzy sets in the same universe. In addition, fuzzy logic works with linguistic variables (i.e. fuzzy sets such as 'height of a male central European' with the values or labels 'short', 'medium', 'tall') which are connected by means of extensions of logical operators such as AND or OR.

Fuzzy inference systems (FIS) are integrated structures consisting of several units that transform information from the input space to the output space by means of fuzzy if-then rules. Fuzzy if-then rules or fuzzy conditional statements are expressions of the form IF A THEN B, where A and B are labels of fuzzy sets characterized by appropriate membership functions [15]. Fuzzy if-then rules may differ only in the types of membership functions in the consequent parts. The consequent part may be defined by either a fuzzy set or a non-fuzzy equation of input variables as given by Takagi and Sugeno [12].

An adaptive neural network (ANN) is a superset of all kinds of feed-forward neural networks with supervised learning capability. It is a network structure consisting of nodes and directional links through which the nodes are connected. Moreover, parts or all of the nodes are adaptive, what means that each output of these nodes depends on the parameter(s) pertaining to a particular node, and the learning rule specifies how these parameters should be changed to minimize a prescribed error measure.

The basic learning rule of ANNs is based on gradient descent and the chain rule.

There are numerous other numerical learning algorithms available, such as those of Gauss-Newton, Levenberg-Marquardt, etc., which could replace the gradient descent method in the hybrid learning algorithm given in Section 2.2. Nevertheless, they have not been considered in this study because the chosen algorithm performed quite well.

ANFIS are feed-forward ANNs which are functionally equivalent to FISs. An FIS is typically designed by defining linguistic input and output variables as well as an inference rule base. However, the resulting system is just an initial guess for an adequate model. Hence, its premise and its consequent parameters have to be tuned based on the given data in order to optimize the system performance. In ANFIS this step is based on a supervised learning algorithm.

2.1 ANFIS Architecture

ANFIS which was proposed by Jang [5], is a multi-layered feed-forward neural network with a supervised learning scheme which is equal to Takagi-Sugeno-Kang (TSK) fuzzy inference system [10, 12].

A common architecture of an ANFIS is shown in Fig. 2, in which a circle indicates a fixed node, whereas a square indicates an adaptive node. For simplicity, we consider two inputs x_1, x_2 and one output f. The ANFIS used in this paper implements a first-order Sugeno fuzzy model. Among many FIS models, the Sugeno fuzzy model is the most widely applied one for its high interpretability and computational efficiency, and built-in optimal and adaptive techniques. For a first order Sugeno fuzzy model, a typical rule set with two fuzzy if-then rules can be expressed as

$$\text{Rule 1: if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } B_1 \text{ then } f_1 = p_1x_1+q_1x_2+r_1 \quad \textbf{(1a)}$$

$$\text{Rule 2: if } x_1 \text{ is } A_2 \text{ and } x_2 \text{ is } B_2 \text{ then } f_2 = p_2x_1+q_2x_2+r_2 \quad \textbf{(1b)}$$

where A_i and B_i are the fuzzy sets in the antecedent, and p_i, q_i and r_i are the design parameters that are determined during the training process. As in Fig. 2, the ANFIS consists of six layers:

Layer 0: Input Layer. It has n nodes where n is the number of inputs to the system.

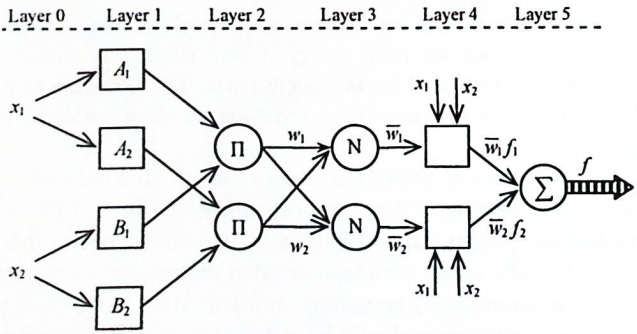


Fig. 2. The ANFIS architecture for a two-input Sugeno fuzzy model with two rules.

Layer 1: Fuzzification Layer. Each node represents a membership value to a linguistic term as a Gaussian function with the mean

$$\mu_A(x) = e^{-\frac{1}{2}\left(\frac{x-c_i}{\sigma_i}\right)^2}. \tag{2}$$

where σ_i and c_i are parameters of the function. These are adaptive parameters. Their values are adapted by use of the back propagation algorithm during the learning stage. As the values of the parameters change, the membership functions of the linguistic term A_i change. These parameters are called premise parameters. In that layer there are $n \times m$ nodes where m is the number of membership functions.

Layer 2: Inference Layer. The nodes in this layer multiply the incoming signals and send the output as shown in the below equation.

$$w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2). \tag{3}$$

The membership values represented by $\mu_{A_i}(x_1)$ and $\mu_{B_i}(x_2)$ are multiplied in order to find the firing strength of a rule where the variable x_1 has linguistic value A_i and x_2 has linguistic value B_i in the antecedent part of Rule i. There are m^n nodes denoting the number of rules in Layer 2. Each node represents the antecedent part of the rule.

Layer 3: Normalization Layer. This layer normalizes the strength of all rules according to the equation

$$\bar{w}_i = \frac{w_i}{\sum_{j=1}^{R} w_j}. \tag{4}$$

where w_i is the firing strength of the ith rule which is computed in Layer 2. Node i computes the ratio of the ith rule's firing strength to the sum of all rules' firing strengths. There are m^n nodes in this layer.

Layer 4: Consequent Layer. The ith node in this layer computes the contribution of the ith rule towards the model output with the function

$$O_i = \bar{w}_i f_i = \bar{w}_i (p_i x_1 + q_i x_2 + r_i). \tag{5}$$

where $\{p_i, q_i, r_i\}$ is the parameter set in the consequent part of the first-order Sugeno fuzzy model.

Layer 5: Output Layer. The function of this layer is the summation of the net outputs of the nodes in Layer 4. The output is computed here as

$$O = \sum_i O_i = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \tag{6}$$

2.2 Estimation of Parameters

The parameters for optimization in an ANFIS are the premise parameters $\{a_i, b_i, c_i\}$, which describe the shape of the MFs, and the consequent parameters $\{p_i, q_i, r_i\}$, which describe the overall output of the system. The basic learning rule of an adaptive network, the back-propagation algorithm [9], which is based on the gradient descent rule, can be successfully applied to estimate these parameters. However, Jang [4] argues that the gradient descent method is generally slow and is likely to get trapped in local minima. Jang has proposed a faster learning algorithm, which combines the gradient descent method and the least squares estimate (LSE) to identify parameters, as described below:

The adaptive network has one output and is assumed to be

$$\text{output} = F(\vec{I}, S). \tag{7}$$

where $\vec{I}$ is the set of input variables and S is the set of parameters. If there exists a function H such that the composite function $H \circ F$ is linear in some of the elements of S, then these elements can be identified by the least squares method. More formally, if the parameter set S can be decomposed into two sets

$$S = S_1 \oplus S_2. \tag{8}$$

(where $\oplus$ represents the direct sum) such that $H \circ F$ is linear in the element S_2, then applying H to Eq. (7), we have

$$H(\text{output}) = H \circ F(\vec{I}, S). \tag{9}$$

which is linear in the elements of S_2. Now given values of elements of S_1, the P training data can be plugged into Eq. (9) to obtain the matrix equation:

$$AX = B. \tag{10}$$

where X is the unknown vector whose elements are parameters in S_2. Let $|S_2| = M$, then the dimensions of A, X and B are $P \times M$, $M \times 1$ and $P \times 1$, respectively. Since P (number of training data pairs) is usually greater than M (number of linear parameters), this is an over-determined problem and generally there is no exact solution to Eq. (10). However, a least square estimation (LSE) of X can be sought that minimizes the squared error $\|AX - B\|^2$.

From the ANFIS architecture presented in Fig. 2 it is observed that given the values of the premise parameters, the overall output can be expressed as linear combinations of consequent parameters. More precisely, the output f can be rewritten as,

$$\begin{aligned} f &= \overline{w}_1 f_1 + \overline{w}_2 f_2 \\ &= (\overline{w}_1 x_1) p_1 + (\overline{w}_1 x_2) q_1 + (\overline{w}_1) r_1 + (\overline{w}_2 x_1) p_2 + (\overline{w}_2 x_2) q_2 + (\overline{w}_2) r_2 \end{aligned} \tag{11}$$

which is linear in the consequent parameters (p_1, q_1, r_1, p_2, q_2 and r_2). As a result, the total number of parameters (S) in an ANFIS can be divided into two such that S_1 = set of premise parameters and S_2 = set of consequent parameters. Consequently the hybrid learning algorithm, which combines the back propagation gradient descent and

least squares method, can be used for an effective search of the optimal parameters of the ANFIS. More specifically, in the forward pass of the hybrid learning algorithm, the node output goes forward until layer 4 and the consequent parameters are identified by the least squares method. In the backward pass, the error signal propagates backwards and the premise parameters are updated by gradient descent. As mentioned earlier, the consequent parameters thus identified are optimal under the condition that the premise parameters are fixed. Accordingly, the hybrid approach converges much faster since it reduces the dimension of the search space of the original back-propagation method. A detailed description of this algorithm can be found in [6].

3 Model Development

Like other neural networks, ANFIS also requires a training process to examine the relationships between each pair of inputs and the corresponding outputs. The sample data applied to this paper are composed of weekly data from the beginning of 1995 through 2004 for a total of 500 observations. The first 80% of the data is training data set, and the last 20% serves as the testing data set whose purpose is to see how well the ANFIS model predicts the corresponding set of data output values. The error criterion of the supervised training is the smaller mean absolute percentage error (MAPE), the smaller root mean squared error (RMSE) and the larger R-squared (R^2).

We have employed MATLAB Fuzzy Logic Toolbox to model the ANFIS structure [2]. The goal is to estimate the West Texas Intermediate (WTI) crude oil price in one week in front, so if $P(t)$ represents the average price of t, the output of both models would be $P(t+1)$ that is the average price at $t+1$. The first model is a technical analysis based (TA-based) model, while in the second one a combination of technical analysis and fundamental analysis is used for prediction of the future oil prices. Here, the latter is called technical-fundamental analysis based (TFA-based) model. These two ANFIS models are developed during the analysis with the corresponding input vectors as follows:

TA-based Model $\quad P(t+1) = f(P(t-3), P(t-2), P(t-1), P(t))$

TFA-based Model $\quad P(t+1) = f(P(t-3), P(t-2), P(t-1), P(t), \Delta COES(t))$

where $\Delta COES(t) = COES(t) - COES(t-1)$ and $COES$(t) corresponds to the US weekly crude oil ending stock at time t.

4 Results

To assess the models' performances, several criteria are used and shown as below.

Root Mean Square Error (RMSE). Evaluates the residual between real and forecasted oil price. This index assumes that larger forecast errors are of greater

importance than smaller ones; hence they are given a more than proportionate penalty

$$\text{RMSE} = \left[\frac{\sum (P_t - O_t)^2}{n}\right]^{0.5}. \tag{14}$$

R-squared (R^2). Represents the

$$R^2 = 1 - \frac{\sum (P_t - O_t)^2}{\sum (P_t - \overline{P}_t)^2}. \tag{12}$$

where P_t is the real price, O_t is the model output, and $\overline{P}_t$ is the average of P_t values.

Mean Absolute Percentage Error (MAPE). Is calculated according to the equation

$$\text{MAPE} = 100 \times \frac{1}{n} \times \sum \frac{|P_t - O_t|}{|P_t|}. \tag{13}$$

The errors of developed models are shown and compared to the errors of a TA-based ANN model in Table 1. TA-based model has the testing MAPE equal to 2.1944.

Table 1. Prediction by ANN, TA-based ANFIS model and TFA-based ANFIS model

	ANN Model	ANFIS Model	
	TA-based	TA-based	TFA-based
R^2	0.7863	0.8293	0.9619
MAPE	4.5729	3.9438	2.4724
RMSE	5.1246	4.2310	2.5736

Fig. 3 illustrates the real data and ANFIS output for testing data in the TA-based model. It is seen that in recent years that the oil market has reached high prices, the model performance is too low and causes a large error in forecasting the future oil prices.

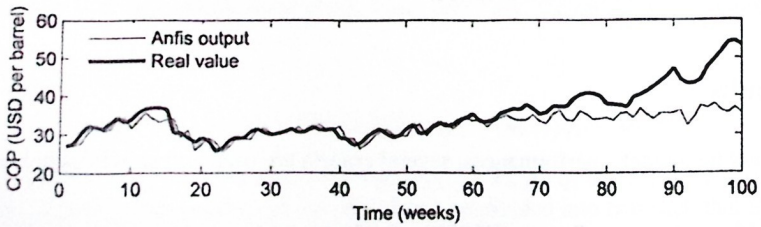


Fig. 3. Real values and outputs for the technical analysis based model.

In Table 1 it is observed that the performances of both ANFIS models are better than ANN model. In addition, we can visualize that adding weekly COES variations cause much better performance in TFA-based ANFIS model. The trends in WTI oil price and predicted price by TFA-based model are seen in Fig. 4.

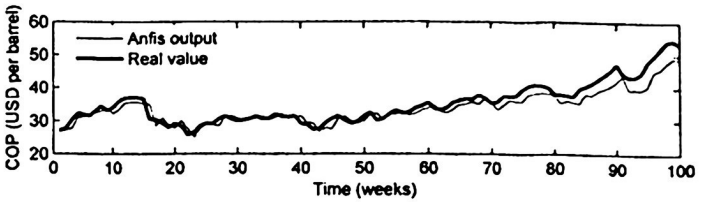


Fig. 4. Real values and outputs for the technical-fundamental analysis based model.

5 Conclusions

In this paper, we propose the use of a fundamental-technical analysis based adaptive neuro-fuzzy inference system (ANFIS) to forecast the future WTI oil prices. We study the role of USA weekly COES in the price variations. Using weekly data sets means that there are 52 observations per year, so for a suitable training, the training and testing data sets cover a wide domain. But the crude oil market follows different behavioral models in different temporal domains. That is the main reason for the low performance of the first model. As Fig. 5 shows, the training data set and testing data set are belonged to two different domains with different structures. In this case, adding weekly COES variations input to the model is a good approach to reach a suitable output.

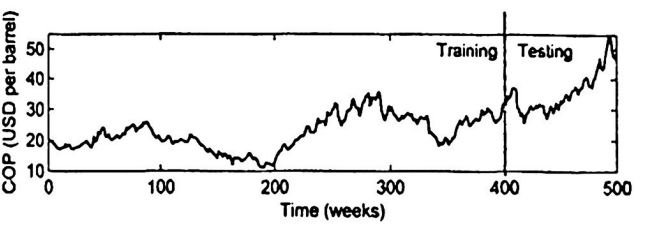


Fig. 5. Training and testing data sets.

The ANFIS preserves the full potential of ANN models, and simplifies the model building process. The results of the study are highly encouraging and suggest that an adaptive neuro-fuzzy approach is viable for modeling the price time series. Though there are other fundamental factors e.g. crude oil imports and exports, refinery operable capacity, percent utilization of refinery operable capacity, US$/Euro exchange rate, temperature and some other social and financial matters that we do not consider them in this paper, but we leave them as a topic for future research. The analyses of the results indicate that the performance of ANFIS models for prediction

of future prices is significantly improved if a combination of technical analysis and fundamental analysis is used.

References

1. Andres, U., Korn, O., Model selection in neural networks, Neural Networks, Vol. 12 (1999) 309-323
2. Fuzzy Logic Toolbox for use with MATLAB®, The Math-Works Inc. USA (2000)
3. Gradojevic, N., Yang, J., Gravelle, T., Neuro-fuzzy decision-making in foreign exchange trading and other applications, Bank of Canada Working Paper, vol. 23 (2000)
4. Jang, J.S., Rule extraction using generalized neural networks, In Proceedings of the fourth IFSA World Congress 4, Volume for Artificial Intelligence, (1991) 82-86
5. Jang, J.S., ANFIS: Adaptive-network-based fuzzy inference systems, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23 (1993) 665-684
6. Jang, J.S., Sun, C.T., Neuro-fuzzy modeling and control, Proceedings IEEE 83, Vol. 3 (1995) 378-406
7. Kutterer, H., Schön, S., Tutorial on fuzzy logic, In: Kutterer H., Carosio A. (eds.) First international symposium on robust statistics and fuzzy techniques in geodesy and GIS, Swiss Federal Institute of Technology Zurich (ETH), Institute of geodesy and Photo grammetry, 12-16 March, (2001) 1-49
8. Pantazopoulos, K.N., Tsoukalas, L.H., Bourbakis, N.G., Brun, M.J., Houstis, E.N., Financial prediction and trading strategies using neuro-fuzzy approaches, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 28, No. 4 (1998) 520-531
9. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (eds.): Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1. MIT Press, Cambridge (1986) 318-362
10. Sugeno, M., Kang, G.T., Structure identification of fuzzy model, Fuzzy Sets and Systems, Vol. 28 (1988) 15-33
11. Sugeno, M., Yasukawa, T., A fuzzy-logic based approach to qualitative modeling, IEEE Transactions On Fuzzy Systems, Vol. 1 (1993) 7-31
12. Takagi, T., Sugeno, M., Fuzzy identification of systems and its application to modeling and control, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 15 (1985) 116-132
13. Wu, I.F., Goo, Y.J., A neuro-fuzzy computing technique for modeling the time series of short-term NT$/US$ exchange rate, The Journal of American Academy of Business, Cambridge, Vol. 7 (2005) 176-181
14. Yager, R.R., A characterization of the extension principle, Fuzzy Sets and Systems, Vol. 18 (1986) 205-217
15. Zadeh, L.A., Fuzzy sets, Information and Control, Vol. 8 (1965) 338-353
16. Zadeh, L.A., (1975) The concept of a linguistic variable and its application to approximate reasoning, part I, (1975) 199-249

Modeling and Control of a Sugar Cane Process

Lázaro Gorostiaga Canepa[1], Yovany González Martínez[2] and José R. Perán[3]

[1,3] Center of Automatization, Robotics and Technologies of the Information and the Manufacture. Spain.
lazgor@cartif.es
[2] Center for Computing Research. Mexico City. Mexico.
yovanygm@sagitario.cic.ipn.mx

Abstract. It is presented the design, modeling and simulation of an automatic control system in sugar cane preparation degree, which makes use of artificial vision techniques for the continuous mensuration of this parameter.This control system comes upon a scientific and technique innovation, rending possible automatic control of the optimum preparation degree, according to differents models of available preparation machines, to later employment of bagasse and others possible changeable interest in the sugar mill.

Key words: modeling, control, robust, vision, automation.

1 Introduction

The preparation station is the first one in a cane sugar factory, and its main objective is to receive the discreet lots of canes coming from the field and to produce a continuous flow of prepared cane and with a optimum preparation degree, in order to feed a mill train or a diffuser, where it is carried out extraction of sugar cane juice.

The main variable to control in the preparation station is therefore the preparation degree, the parameter determining quality of the elaborated product, that is to say the prepared sugar cane.

Modeling and automatic control of preparation degree [1], [2] is a scientific, technological and economic problem, taking into account that in all over the world doesn't exists an automatic control of this parameter and that manual control is extremely inefficient. This can be verified by simple observation of this station operation in different sugar mills. So we can appreciate different preparation degrees in these stations as son as frequently it takes place stops of preparation machines caused by overloads and obstructions.

The preparation degree is a measure in per cent of open cells, and its juice can be extracted by compression in the mills or by diffusion in a diffuser. It can also be a measure of type and size of the particles (also called fineness) that are produced in dividing into pieces and breaking up sugar cane stalks in preparation machines.

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 171-177

Two basic objectives are attended in the preparation station:

To increase juice extraction in mills train or in diffusers when sugar cane stalks fibrous structure are broken and juice cells are free. To increase tandem processing capacity increasing volumetric density of feed sugar cane mattress is the objective.

According to [4] "The entire sugar canes more or less entangled have a low volumetric density because of the great proportion of air space between sugar canes stalks. But when they have been cut in the preparatory plant, we have smaller pieces rather uniforms, and consequently with smaller empty volume. In such a case, the pressed sugar canes in the mills are bigger in proportion

Hugot [3] mentions the following figures about volumetric density: 8-10 lb./cu.ft. for entire sugar canes more or less tangled and 15-20 lb./cu.ft. For those grounded in the mills.

Depending on several factors we can have an optimum preparation. A very fine preparation (greatly small sugar cane particles) causes a degradation in mills process and rend more difficult burnt of the bagasse in the furnaces. According to this some experts recommend that the preparation were as fine as possible.

Jenkins [4] recommends that the preparation should be fixed in a average value, so that optimum results weren't obtained with not one sugar cane types. In such cases another easy means for to vary preparation degree in relation to sugar cane would be very attractive.

There are two generic types of preparation, one with long and narrow pieces and another with particles more or less uniforms.

Different laboratory analytic techniques have been used in measuring the preparation degree, among those we have: sieving, volumetric density, energy consumption, expert visual judgement and others. In commercial practice, visual observation is the one commonly used. Although different observers are generally congenial with comparative evaluation of fineness for a certain number of samples; this method can be applied only if we dispose of samples series as a reference point.

The preparation area is basically conformed by cane conveyors and preparation machines, such as levellers, sugar cane knives and shredders.

Usually, sugar cane bundles are directly discharged from railroad boxcars or trucks to a sugar cane conveyor or to a lateral table, which feeds to a second sugar cane conveyor where are located the preparation machines. As a study case has been selected a plant with two sugar cane knives situated in the second conveyor. Fig 1.shows a scheme of this plant and the designed control system.

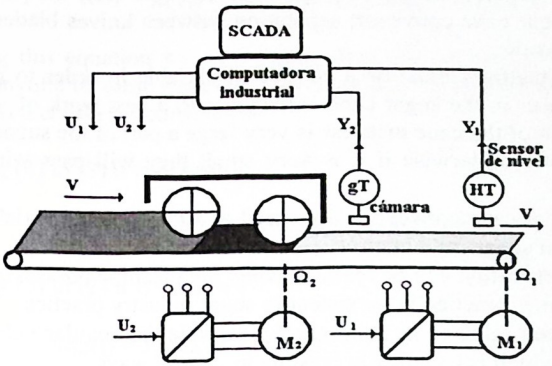


Fig. 1: Lateral view of a cane conveyor with two cane knives.

The knife blades are removable, so that they can be easily taken off for resharpening or for replacing.

The sugar cane knives rotate with angular velocities between 400 and 800 RPM. This machine breaks the sugar canes in small pieces, forming a more condensed sugar cane mattress to make more easy juice extraction in the mills or in the diffuser. When two sets of sugar cane knives are used, the first one (leveller knives) works with a wide separation adjustment (high clearance) allowing to pass a high proportion of uncut sugar canes. The second knife set (cutting knives) completes this work and is used as a very low clearance.

The authors are developing a project of advanced control for the sugar cane preparation plant ,at the time being it has been demonstrated the necessity and convenience in designing and setting up an automatic control of the preparation degree. Therefore it is necessary to obtain a model of the process

2. Modeling in Preparation Process

The preparation degree (g) is the output of the second preparation machine and it depends on the following factors:

- Machine angular velocity.
- Hardness or resistance of sugar canes to be cut, which at once depends of their variety, maturity and freshness.
- Sugar cane flow (F) to be processed.
- Separation between knives blades and sugar cane conveyor level work. Edge of sugar cane knives, which decreases progressively.
- preparation degree apported by the first machine.

In order to control the preparation degree it can be used as manipulated variables:

angular velocity of machine, height of sugar cane mattress (H), the power of machine, velocity of sugar cane conveyor, separation between knives blades and sugar cane conveyor level work .

The height of the mattress must be a controlled variable in order to guarantee a mass automatic balance in the sugar cane conveyor and a best work of preparation machines. If the height of the cane mattress is very large a part of the sugar canes will pass besides the knives; otherwise if it is very small they will pass without being processed.

The velocity of sugar cane conveyor (V) is used as a manipulated variable in order to control the height of sugar cane mattress (H) in this one.

Manipulation of separation can be realized when the machine possesses an adjustable bar-anvil, and this is practically inexistent in sugar industry practice.

Therefore it has been selected as manipulated variable the angular velocity of the preparation machine which is technically feasible in several ways.

Considering a fixed separation, constant height of sugar cane mattress (H) for an automatic regulation loop, sugar cane knives replaced with a certain periodicity and using incremental variables, the following expressions can be obtained for calculation of variation in preparation degree.

$$g(t) = \lambda_1 * \Omega(t) - \lambda_2 * V(t) - \lambda_3 * D(t) - \lambda_4 * \rho(t), \quad \lambda_i > 0 \tag{1}$$

As it is observed, perturbations that affect the preparation degree are sugar cane conveyor velocity

(V), sugar cane hardness (D) and volumetric density of sugar cane mattress. An increment of anyone of them causes a reduction of the preparation degree and vice versa.

The preparation machines can be moved by electric motors, hydraulic motors and steam motors. Whichever it is the type that is used, the following equation can settle down:

$$M_m(t) = M_a(t) \tag{2}$$

Where Mm(t) y Ma(t) are respectively the motor and the antagonistic moments. At once the antagonistic moment is calculated by equation (3)

$$M_a(t) = J \times d\Omega/dt + B \times \Omega(t) + M_c(t) \tag{3}$$

Where J= Jmotor + Jmachine, B is the coefficient of viscous damping and Mc(t) is the load moment opposed by the sugar canes mattress while they are cut and broken in pieces.

In non linear form the load moment depends of sugar cane hardness, height of the mattress, velocity in crossing the machine, knives edge and angular velocity.

Under previous assumptions, considering small variations around the nominal operating point, and using deviation variables we can write:

$$M_c(t) = \alpha_1 \times \Omega(t) + \alpha_2 \times V(t) + \alpha_3 \times D(t) \quad (4)$$

Applying this equation we can obtain variations in load moment opposed by the knives. If anyone of these variables increases very little over nominal value the load moment will increase proportionally.

Substituting (4) in (3) it is obtained

$$M_a(t) = J \times d\Omega/dt + B \times \Omega(t) + \alpha_1 \times \Omega(t) + \alpha_2 \times V(t) + \alpha_3 \times D(t) \quad (5)$$

Substituting (5) in (2) it is obtained

$$M_m(t) = J \times d\Omega/dt + B \times \Omega(t) + \alpha_1 \times \Omega(t) + \alpha_2 \times V(t) + \alpha_3 \times D(t) \quad (6)$$

Taking the Laplace transform on both sides of equation (6) with initial conditions assumed to be zero we have:

$$\Omega(S) = \frac{1}{JS + B + \alpha_1} \times \left(M_m(S) - \alpha_2 \times V(S) - \alpha_3 \times D(S)\right) \quad (7)$$

Equations (1) and (7) represent a linear model of the process, in the environment of nominal operating point. The coefficients of the model depend on the specific plant and some of them are difficult to obtain.

This previous analysis shows that a plant conformed by a frequency converter, motor, machine and the conveyor can be represented by a first order system with a pure retard whose equation is:

$$\frac{g(S)}{U(S)} = \frac{K \cdot e^{-LS}}{1 + TS} \quad (8)$$

The plant gain (K) is uncertain due to the gradual loss of knives edge, between a substitution of the blades and the following one.

Using experimental methods the following values were calculated:
K_{max} = 5,0 (recently sharp knives)
K_{min} = 1.5, (blades with a week of use)
K_{nom} = 2.5 (nominal value)

The total delay of model L is the addition of real delay in the plant and apparent delay due to considerate a high order system as a first order one. As we will show later, the former is greater than the apparent one.

The real or pure delay is the time delay (L) introduced by the sugar cane conveyor. It depends of distance " d " between video camera and the preparation machine output, linear velocity (v) of sugar cane conveyor which is the manipulated variable in the regulation loop of the sugar cane mattress height (D)

According to this it will be advisable to place video camera near the preparation machine in order to reduce time delay.

In this case, we have estimated in 3 meters the minimum distance for to place video camera and the maximum range of sugar cane conveyor velocity from 1 to 6 m/min. So we will have that time delay will vary between 30 and 180 seconds.

Finally, using experimental techniques we have estimated apparent time constant of the model, T = 40 seconds, this parameter is associated to dynamic of frequency converter and motor.

Conclusively, it is a process with a dominant delay and as an aggravating circustance, it is variable in a wide range because of its difficult dynamics.

3. Analysis of Different Control Algorithms.

The control system components are an industrial computer, an artificial vision kit and a frequency converter varying velocity of squirrel-cage motor

It was decided to evaluate different control configurations, considering at the beginning the simplest and common, that is to say, a PID standard controller syntonized by Ziegler-Nichols classic method.

Delay (Lm) introduced by processing images techniques of prosecution of images for measurement of preparation degree is supposed constant and equal to 5 seconds.

Using simulation techniques the system was carry to the limit of stability, with integral and derivative actions disabled; the following values were obtained for critical gain and critical period: Kpcrít = 0. 55 y Pcrít = 250 s.

Using closed loop Ziegler-Nichols formulas; it was determined the controller's parameters: Kp=0.33, Ti=175 y Td =31.25.

After Tuning of the controller, the system was simulated for plant nominal values and it was obtained the following indicators: 5 % overshoot, time response equal to 800s approximately and ITAE= 215 for 1500 s in simulation.

Later on, parameters of the plant were varied in correspondence to their nominal values in order to evaluate robustness of the system. It was observed that for nominal delay of 100s with a gain of 3.8 (smaller than the maximum gain) the system is already in the limit of stability.

Therefore, we can conclude that this alternative of control is not adequate in this process of difficult dynamics, with dominant delay and also variable.

As a second alternative, it was used the method proposed in [1] to adjust conventional controllers PID in robust form. Supposing d=0.5 the controller's parameters are fixed in the following way:

$$K_p = \frac{1}{K} \bullet \frac{T + \frac{L}{2}}{L(D+1)} = 0.24 \quad T_i = T + \frac{L}{2} = 90 \qquad T_d = \frac{T \bullet L}{2T + L} = 22.22 \quad (9)$$

Like in the previous case, the system was simulated for plant nominal values, and it was obtained the following indicators: 3 % maximum overshoot, time response of 500 s approximately, and ITAE= 160 for 1500 s in simulation.

Later on, varying parameters of the plant as in the other case in order to evaluate robustness of the system, it was observed an improvement in the system behavior but with an insufficient value of robustness for controlling the plant. For example, with a process gain k= 5 and a time delay L = 125 seconds the system is already in the stability limit.

The design of a robust controller for the regulation of the sugar cane mattress level is presented in [5].

4. Conclusions

The hardware components of the system are a computer vision kit, a data acquisition card, an industrial control computer and a frequency converter. The PID controller is in the industrial control computer.

The control system uses artificial vision and processing patterns in order to measure preparation degree on line and continually.

The video camera will be placed as near as possible to the preparation machine to reduce to a minimum time delay introduced by the sugar cane conveyor.

This automatic regulation loop guarantees optimum preparation degree required by the sugar mill and it also prevents overload stops in preparation machines. The control system can be easily extendrd to others preparation plants different from that analyzed in this paper.

The dynamics of the preparation process is characterized by its great uncertainty, reason why the controller PID syntonized by classic methods cannot guarantee the demanded benefits.

Consequently it is necessary to use advanced algorithms for synthesis of a controller with sufficient robustness.

5. References

1. Dario, E.: Robust adjustment to conventional controllers. Form simple to apply it. Comparison with the classic methods and the use of predicting. Congress of the Association Argentina of automatic Control, 1998
2. Gorostiaga, L.: Dynamics and Sugar Technological Process control. Doctoral thesis. 1998.
3. Hugot, E.: Manual for Sugar Engineers. 1980.
4. Jenkins, G.: Introduction to the technology of the cane sugar. 1988.
5. Loma, M.; Baeyens, E.; Gorostiaga, L. and Peran, J.: Robust control of an elevating cane mat. Computer science 2000. Havana, Cuba. 2000.

Implementing RTSJ-optimized Java Processor for Real-Time Applications

Zhilei Chai[1], Wenke Zhao[2], Wenbo Xu[1]

[1] Center of Intelligent and High Performance Computing, School of Information Technology, Southern Yangtze University, 214122, Wuxi, China
zlchai@fudan.edu.cn
xwb@sytu.edu.cn
[2] Department of Computer Science and Engineering, Fudan University
200433, Shanghai, China
052021110@fudan.edu.cn

Abstract. Due to the preeminent work of the Real-Time Specification for Java (RTSJ), Java is increasingly expected to become the leading programming language in embedded real-time systems. To provide an efficient Java platform suitable for real-time applications, a hard real-time Java processor (HRTEJ) can execute Java bytecode directly is proposed in this paper. It efficiently supports mechanisms specified by the RTSJ and offers a simpler programming model through ameliorating the scoped memory of the RTSJ. The worst case execution time (WCET) of this processor is predictable by employing the optimization method proposed in our previous work[1], in which all the processing interfering predictability is processed before bytecode execution. Further advantage of this method is to make the implementation of the processor straightforward and suited to a low-cost FPGA chip.

1 Introduction

The complexity of embedded real-time systems keeps growing and novel methods and tools are required. With the advantages as an object-oriented and concurrent programming language, and with the occurrence of the real-time specification for Java (RTSJ)[2], Java is increasingly expected to become the leading programming language in embedded real-time systems.

Currently, to provide an efficient Java platform suitable for real-time applications, many different implementing methods are tried. real-time Java platforms can be generally classified as follows according to their implementing methods: 1). **Interpreter,** in this manner, Java platform is an interpreting program running on the operating system that can interpret and execute java bytecode under the assistance of the OS. RJVM[3], JTime[4], and Mackinac[5] all belong to this way. 2). **Ahead-of-Time** Compiler, in this manner, Java bytecode is compiled into native code or an intermediate language (e.g. C) in advance. Anders Nilsson et al[6] employed this method to insure real-time performance of the Java applications. In their implementation, C is taken as the intermediate language. 3). **Java Processor,** in this manner, the Java plat-

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 179-188

form is implemented directly on silicon. It not only avoids the overhead of translation of the bytecode to another processor's native language, but also provides special support for Java runtime features such as stack processing, multithreading etc. aJile[7], FemtoJava[8] and JOP[9] are all real-time Java processors implemented in the FPGA.

Comparing with other implementing techniques, Java processor is preferable in embedded systems for its high execution efficiency and low memory footprint. Therefore, it becomes popular in implementing embedded real-time Java platforms. Presently, FemtoJava and JOP do not put emphasis on supporting the RTSJ. aJile systems announces the RTSJ will be supported on top of the aJ-80 and aJ-100 chips.

In this paper, we propose a simple Java processor suitable for embedded real-time applications and it is dedicated to supporting RTSJ features more efficiently. This processor implements RTSJ mechanisms effectively and offers a simpler programming model through ameliorating the scoped memory of the RTSJ. According to the optimization method proposed in work[1], all the processing interfering predictability is completed before bytecode execution, thus, this processor is simple to implement and its worst case execution time (WCET) is more predictable.

2 RTSJ related concepts

RTSJ is a specification for extending the Java language specification[10] and the Java virtual machine specification[11], and providing an application programming interface(API) that will enable the creation, verification, analysis, execution, and management of Java threads whose correctness conditions include timeliness constraints. It contains some enhanced areas such as thread scheduling and dispatching, synchronization and resource sharing, asynchronous event handling, asynchronous transfer of control, asynchronous thread termination, memory management, and physical memory access. The details of the RTSJ can be found in [2]. In this paper, we put emphasis upon the special support for the RTSJ mechanisms in our processor, the methods to guarantee its real-time performance is discussed as well.

3 Implementation of the HRTEJ processor

As described in [1], to guarantee the real-time performance of the processor, standard Java class files was processed by the *CConverter* (the program we designed to preprocess the Java class file) before being executed on the processor. During this phase, all the classes needed in the application are loaded, linked and transformed into the binary representation shown in Fig.1 that can be executed directly on the processor. Once power on, the HRTEJ processor starts to execute *Initial* code to do system initialization according to the *Initial* parameters. Then, it enters main thread creating code and executes bytecode from there step by step.

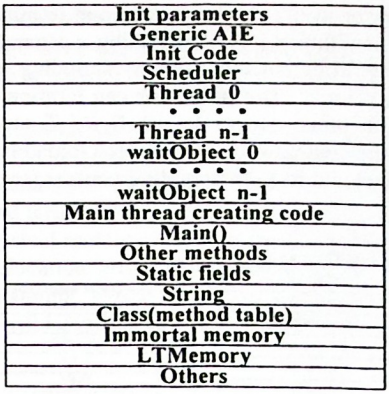


Fig.1. Memory layout of the binary representation.

3.1 Thread management in the HRTEJ processor

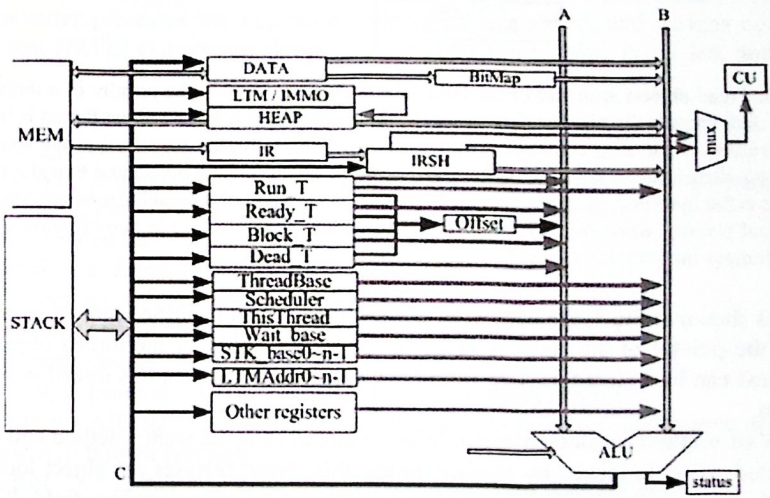


Fig.2. Thread related architecture of the HRTEJ processor. *Run_T*, *Ready_T*, *Block_T* and *Dead_T* are the n-bit (n is the width of the data path) registers to record the queues of threads which are running, ready, blocked and dead respectively. A thread can be put into a queue by setting corresponding bit of that register to '1' according to this thread's priority. *ThisThread* records the object reference of current running thread. *Wait_Base* is the base address of the static fields *WaitObject0~n-1* in Fig.1. *STK_base0*~n-1 is the base address of the stack for each thread. *LTMAddr0*~n-1 is the base address of the LTMemory memory for each thread.

The HRTEJ processor can support n (n is the width of the data path) threads at most with unique priority from 0 to n-1 (0 is the highest priority).These threads (several

threads are kept for processor itself) can be created and terminated dynamically during execution. Creating a new thread just like creating a general object, but the object reference of this thread should be put into the corresponding static field *Thread0~n-1* according to this thread's priority. The processor can terminate a thread by moving the corresponding '1' from other queues to the *Dead_T* according to its priority. When scheduling occurs, the context of the preempted thread is saved at the top of its own stack. The thread object and its corresponding context (registers in the processor) are shown in Fig.3.

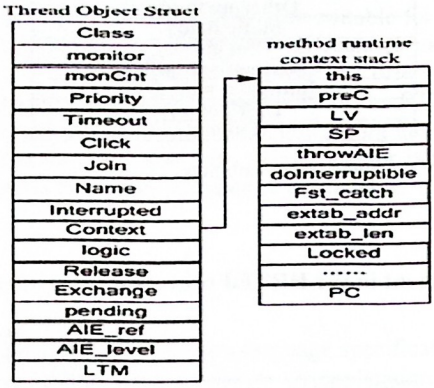


Fig.3. Tread objects structure of the HRTEJ processor. *Priority* is the priority of a thread (0 to n-1). *Join* records the object reference of the thread to wake up when current thread is finished. *Interrupted* is the flag compatible with the conventional thread to record if the *interrupt*() method of this thread is called. *Context* is the stack pointer of the preempted thread's context. *Logic* is the bytecode sequence executed by the *run*() method of a thread. *Exchange* records the original priority when priority inheritance avoidance is taken. *LTM* is current address of the LTMemory memory for a thread.

As shown in Fig.3, the context of a preempted thread is pushed into its own stack, and the pointer of the stack is saved in the field *Context* of the thread object. The context can be restored starting from the *Context* pointer when this thread is selected again.

Wait method implementation: When a thread calls the wait() method and the requested object is locked by another thread, this thread releases the object locked by itself and records the requested object reference in corresponding field *WaitObject0~n-1* according to its priority and blocks itself. This static field will be checked when another thread calls the *notify*() method.

Join method implementation: Using the instance field *join* to record the object reference of the thread to wake up when current thread is finished.

Priority inheritance implementation: If a thread wants to enter a synchronized block which another thread with a lower priority is in, then the priority inheritance must be taken. In HRTEJ, a simple method to implement the priority inheritance is adopted. The scheduler checks the field *Exchange* of the thread owning the shared object, if the priority inheritance has been taken (*Exchange* != -1), the higher priority

will be assigned to this thread directly. Otherwise, the original priority of this thread is saved in its *Exchange* field, then, the higher priority is assigned to it. When this thread releases the locked object, it takes the original priority back again from *Exchange*.

As discussed above, special hardware is used in the HRTEJ processor to ensure the predictability of WCET. The clock cycles of thread scheduling, dispatching, and other thread related mechanisms are all predictable in the HRTEJ processor.

Asynchronous thread termination is processed in the same way with asynchronous transfer of control. Physical memory access is not supported in HRTEJ to conceal the details of memory allocation from Java programmers. The special hardware in HRTEJ for thread scheduling, synchronization, asynchronous transfer of control and memory management will be introduced in following sections.

3.2 Implementing ATC in the HRTEJ processor

1 ATC preprocessing by the CConverter

CConverter reads standard Java class file and converts all of the methods into a binary format. Attributes of each method are stored in fields with a determinate location.

CConverter processes the exception table of every method, and assigns correct values that HRTEJ can process directly to the items *extab_addr*, *extab_len*, and *extab_*item.

2 ATC triggered by target.interrupt() or target.aie.fire()

The ATC is triggered by invoking method *interrupt()* or *aie.fire()* in a thread.

cur_level, *cur_AIE* is variants defined in *interrupt()* or *aie.fire()*.

ATC related fields of each thread shown in Fig.4.

Pending: it shows there is an AIE in action.

AIE_ref: the reference of the received AIE.

AIE_level: the method invocation level of the received AIE.

```
if(target.pending == 0){                    //no AIE in pending
    target.pending = 1;          //target thread is marked pending
    target.AIE_ref = cur_AIE; //record the object's reference for generic or specific AIE thrown currently
}else{       //other AIE in pending, replacement rules must be taken
    if(cur_AIE == GenericAIE){//generic AIE has a higher priority
            target.AIE_ref = cur_AIE;
    }elseif(cur_AIE!=GenericAIE&&target.AIE_ref!=GenericAIE){
        if(target.AIE_level>cur_level){//record the AIE with higher priority.
            target.AIE_ref = cur_AIE;
        }
    }
}
```

3 ATC processed by the scheduler

ATC related registers in HRTEJ:
throwAIE: it denotes whether the current method has a throws AIE clause.
doInterruptible: denote if there are catch AIE or its superclass, finally clauses in this method.
Fst_catch: record the first reference of method context be used to process the AIE.
Locked: denoting if this method is a synchronized method or not.

When the target thread of the thrown AIE is scheduled again, the scheduler will process it based on three different conditions. **When** the target thread is in an AI-deferred section, the scheduler restores its context and executes it as a normal thread. **When** the target thread is in an AI section and it is the run() method of interface *doInterruptible*, the scheduler pops the stack frame of run() method and restores the context of method interruptAction() to handle the AIE. **When** the target thread is in an AI section and it is in other methods instead of run() method of interface *doInterruptible*, the scheduler will restore context from the register *fst_catch* to handle the AIE.

3.3 Memory management in the HRTEJ processor

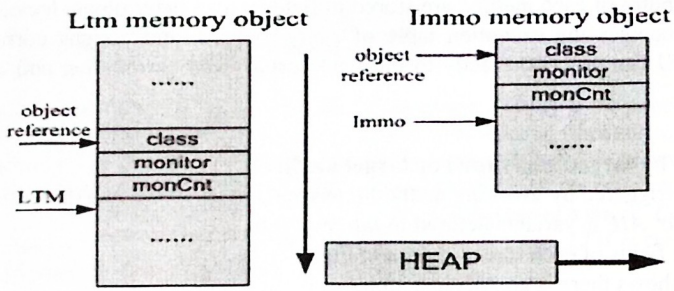


Fig.4. Memory management of the HRTEJ processor. *LTM* is current pointer to allocate space in the *LTMemory* of the running thread. *Immo* is current pointer to allocate space in the Immortal Memory. *HEAP* points to *LTM or IMMO* to concretely complete an allocation.

Memory management policy in the HRTEJ processor:

```
New LTMemory / ImmortalMemory:
        class(classAddr)=>mem[LTM/Immo++];
        monitor(0) => mem[LTM/Immo ++];
        monCnt (0) => mem[LTM/Immo ++];
        return the LTMemory/ImmortalMemory object reference(LTM/Immo-3);

LTMemory.enter• • //'1'denotes LTM; '0'denotes IMMO;
        '1' => MEMType; LTM => HEAP;
LTMemory.exit()
        this => LTM:        //go back to last LTM scope

ImmortalMemory.enter()
        '0' => MEMType; Immo => HEAP;
```

```
New GeneralObject:
        HEAP+ObjectSize => HEAP
        if(MEMType == '0')          HEAP => Immo;    //update Immo
        Else                        HEAP => LTM;     //update LTM
```

To avoid the interference of garbage collector and keep the efficiency about memory space, the LTMemory is provided in the HRTEJ processor. In standard RTSJ, the memory size must be specified by the programmer to use the LTMemory. It makes the programming model a little tricky to Java programmers. As described above, in the HRTEJ processor, an API *exit*() is offered to avoiding the programmer to specify the memory size for a LTMemory. This programming model is more maneuverable for the Java programmer. Moreover, without shared LTMemory between threads and based on the operation policy above, the memory management WCET of the HRTEJ processor is predictable.

4 Evaluation and discussion

The HRTEJ processor is implemented in experimental platform FD-MCES which provides a FPGA chip XC2S150-PQ208 and some debugging conveniences. Through the monitoring software FD-uniDbugger, the bytecode execution on top of HRTEJ can be traced single cycle. Due to the constraints of the experimental platform, the current version of HRTEJ is 16-bit and about 100 instructions implemented with several extended instructions. The memory used is 32Kx16 with 0.1us latency, so, read and write without cache by HRTEJ with 8M frequency can be completed in one cycle that simplifies the WCET analysis.

Table 1. Clock cycles of bytecode execution time

	HRTEJ	HRTEJ	JOP	JOP
iload	2	2+r	2	2
iadd	1	1	1	1
iinc	8	7+r	11	11
ldc	8	6+2*r	4	3+r
if_icmplt taken	10	9+r	6	6
if_icmplt not taken	10	9+r	6	6
getfield	7	5+2*r	12	10+2*r
putfield	7	5+w+r	15	13+r+w
getstatic	8	6+2*r	6	4+2*r
putstatic	8	6+w+r	7	5+r+w
iaload	2	2+r	21	19+2*r
invoke	43	34+9*r	82	78+4*r+b
invoke static	39	29+10*r	61	58+3*r+b
invoke interface	n/a	n/a	90	84+6*r+b
dup	2	2	1	1
new	12	12	Java	Java
iconst_x	2	2	1	1
aconst_null	2	2	1	1
astore_x	3	3	1	1
aload_x	3	3	1	1
return	20	20	14	13+r+b
ireturn	22	22	16	15+r+b
goto	5	4+r	4	4
bipush	4	4+r	2	2
pop	1	1	1	1
istore	2	2+r	2	2
istore_x	3	3	1	1

Table 1 shows some bytecode execution time in cycles implemented by the JOP and the HRTEJ processor. r denotes the time to complete a memory read and w denotes a memory write when the bytecode needs access the memory. b is the bytecode loading time when a cache miss happened in JOP. There is no instruction/method cache implemented in HRTEJ till now, so b is not used for denoting its bytecode execution cycles. The column 1 and 3 show the bytecode execution time assuming that r=w=1 and b=0. Obviously, the JOP with 100M frequency has a much higher performance than HRTEJ. Currently, the HRTEJ processor puts more emphases on the predictability than the performance. The performance will be considered carefully at the next step.

Estimating the WCET of tasks is essential for designing and verifying a real-time system. Because the measured execution times of tasks is sensitive to their inputs. As a rule, static analysis is a necessary method for hard real-time systems. Therefore, the WCET of the example shown in appendix is static analyzed to demonstrate the real-time performance of the HRTEJ processor.

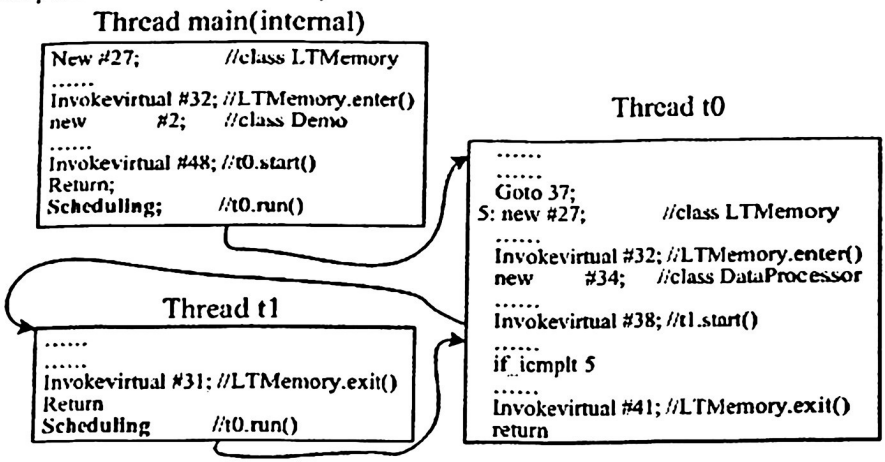


Fig.5. Thread scheduling of example *Demo.java* in appendix. The bytecode compiled from *Demo.java* can be mainly partitioned as 3 parts. In each part, the WCET of the general bytecode listed in Table 2 is predictable. For thread *t0*, there is a finite loop. Its WCET can be calculated as *100*WCET (general code + LTMemory + start() + Scheduling)*. Method *start()* sets the created thread into queue *Ready_T* and wait scheduling. *Scheduling* denotes the WCET of the scheduler execution when a thread scheduling happens. The WCET of the *LTMemory* operation is predictable as shown in Fig.4. Thus, just the WCET of the method *start()* and *Scheduling* are analyzed in detail below.

The process of method *t.start()* and its WCET:

Disable the interrupt;	(1	cycle)
Save the PC for current thread;	(1	cycle)
Put current thread into Ready_T;	(2	cycles)
Jump to the scheduler;	(1	cycle)

The process of *Scheduling* and its WCET:

Disable the interrupt;	(1	cycle)
Save the context of the preempted thread;	(17	cycles)
Move corresponding '1' from *Run_T* to *Ready_T*;	(2	cycles)
Move the leftmost '1' to the corresponding bit in *Run_T*;	(2	cycles)
Save the thread reference to *ThisThread* register;	(3	cycles)
Restore the context for the thread corresponding to the leftmost '1';	(16	cycles)

From discussed above, the WCET of the method start() and scheduling is also predicted. So, the real-time performance of the whole application can be guaranteed. Furthermore, in appendix, the efficiency of the LTMemory is illustrated clearly. The maximal allocation of the LTMemory space in this example is S(t0)+S(t1) instead of S(t0)+100*S(t1). S(t) denotes the space of thread t.

Another advantage learned from this example is that the programming model of the LTMemory is very simple. Java programmers just need creating and entering a LTMemory space to use it instead of denoting its memory size.

5 Conclusions

In order to employ Java programming language to improve development efficiency, security and robustness of real-time systems, many different real-time Java platforms were proposed. Due to high execution efficiency, low memory footprint and low power consumption, Java processor is preferable for embedded real-time systems. In this paper, a hard real-time Java processor based on the RTSJ is implemented. This processor provides some special supports for the mechanisms specified in the RTSJ such as ATC, scoped memory etc. Because most of the operations are preprocessed by the CConverter, this processor is simple to implement and suitable for low cost real-time systems.

References

1. Z. L. Chai, Z. Q. Tang. L. M. Wang. and S. L. Tu, "An Effective Instruction Optimization Method for Embedded Real-Time Java Processor," 2005 International Conference Parallel Processing Workshops, Oslo, Norway, pp. 225-231, 2005.
2. G. Bollela, J. Gosling. B. Brosgol, P. Dibble, S. Furr, D. Hardin, and M. Trunbull, "The Real-Time Specification for Java," Addison Wesley, vol. 1st edition, 2000.
3. H. Cai and A. Wellings, "Supporting mixed criticality applications in a ravenscar-java environment," ON THE MOVE TO MEANINGFUL INTERNET SYSTEMS 2004: OTM 2004 WORKSHOPS, PROCEEDINGS, vol. 3292, pp. 278-291, 2004.
4. "TimeSys JTime® 1.0 RTSJ Extensions Reference Pages," available at http://students.cs.tamu.edu/jtime_guide.pdf.
5. G. Bollella, B. Delsart, R. Guider, C. Lizzi, and F. Parain, "Mackinac: making HotSpot/spl trade/ real-time," presented at Eighth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing, 2005. ISORC 2005., 2005.

6. A. Nilsson and S. G. Robertz, "On real-time performance of ahead-of-time compiled Java," presented at Eighth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing, 2005. ISORC 2005., 2005.
7. D. S. Hardin, "aJile Systems: Low-Power Direct-Execution JavaTM Microprocessors for Real-Time and Networked Embedded Applications," Available at http://www.ajile.com/Documents/aJile-white-paper.pdf.
8. S. A. Ito, L. Carro, and R. P. Jacobi, "Making Java work for microcontroller applications," Design & Test of Computers, IEEE, vol. 18, pp. 100 - 110, 2001.
9. M. Schoeberl., "JOP: A Java Optimized Processor for Embedded Real-Time Systems," Phd dissertation, http://www.jopdesign.com, 2005.
10. J. Gosling, B. Joy, G. Steele, and G. Bracha, "The Java Language Specification Second Edition," 2000.
11. T. Lindholm and F. Yellin, "The Java Virtual Machine Specification, 2nd edition," Addison Wesley, 1999.

Appendix

Demo.java

```
import javax.realtime.*;
class DataProcessor extends NoHeapRealtimeThread{
        int data=0;
        public DataProcessor(int priority){
                super(priority);
        }
        public void run(){
                System.out.println("Processing in DataProcessor");
                Demo.ltm1.exit();
        }
}
public class Demo extends NoHeapRealtimeThread{
        public static LTMemory ltm = null;
        public static LTMemory ltm1 = null;
        public Demo(int priority){
                super(priority);
        }
        public void run(){
                for(int i=0; i<100; i++){
                        ltm1 = new LTMemory();
                        ltm1.enter();
                        DataProcessor t1 = new DataProcessor(3);
                        t1.start();
                } ltm.exit();
        }
        public static void main(String[] args){
                ltm = new LTMemory();
                ltm.enter();
                Demo t0 = new Demo(5);
                t0.start();
        }
}
```

Estimation of Symmetric Positive Definite of Matrices using Evolution Strategy for Differential Games

Hector Vargas[1], Vittorio Zanella[1], Vladimir Alexandrov[2], Mónica López[3] and Lorna Rosas[1]

[1] Autonomous Popular University of Puebla State, Computer Engineering, Santiago 1103, Puebla,Pue. 72160, Mexico
{hectorsimon.varga, vittorio.zanella, lornaveronica.rosas}@upaep.mx
http://www.upaep.mx

[2] Moscow State University, Mechanics and Mathematics, Glavnoe Zdanie, Leninskiye Gory, Moscow, 119899, Russian Federation
valexv@mech.math.msu.su
http://www.msu.ru

[3] Autonomous University of Puebla, Computer Engineering, University City, Puebla,Pue. 72570, Mexico
moniarchundia@cs.buap.mx
http://www.cs.buap.mx

Abstract. Estimation of symmetric positive definite of matrices is required when solving a variety of control problems including differential game theory, robotic control, smart structure control, and intelligent control. In differential game theory, the minimization or maximization of a functional is carried out, which is provided by means of the criterion of a quadratic integral whose parameters are: the coordinates of the dynamic system, the control information and the symmetric positive definite class, and the accuracy of its estimate directly affect control performance. These symmetric positive definite of matrices are obtained based on the engineer's experience on one dynamic system in particular. It is clear that the *good* choice of matrices, in an enclosed set, should improve the criterion's objective. In order to automate the choice of symmetric positive definite of matrices, we propose an algorithm that uses evolution strategy. The important result is obtained by means of two propositions which show that the operations of recombination and mutation preserve the symmetric positive definite properties of the matrices, thereby improving the computational complexity of said algorithm.

1 Introduction

In the theory of differential games [1], the problem of computational testing can be built and solved [2] by applying the concepts of minimax and maximin, that is, by obtaining an algorithm that evaluates and corrects the control strategy in order to stabilize a controllable dynamic system. During the process of building the test algorithm, the expert engineer must choose, based on his experience, confidence and

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 189-197

intuition, the positive definite matrices that obey the necessities of the dynamic system, taking into consideration its physical nature. The present article presents a way of automating the experience of the human expert by using evolution strategy (ES) [3], [4], [6]. The important result is achieved through two proposals which refer to the fact that the recombination and mutation operations preserve the positive definite properties of the matrices, thereby improving the computational complexity of said algorithm.

2 Statement Problem

We consider the following ([2],[7],[8]) for the computer tests: we suppose that an unknown, control algorithm exists, from which we are only able to know its output $u_1 = u(t,x)$, and we furthermore want to organize a computer testing system in order to know how good this algorithm is. In other words, if we wanted to evaluate the said algorithm, how could we obtain an excellent evaluation with which we could compare the obtained evaluation with the said algorithm? Under the same conditions as in the minimax case [2], the maximin problem can be proposed:

$$\inf_{u_1 \in U} \Im\left(x(t_0), p, u_1\right) \rightarrow \sup_{\substack{x(t_0)\in X_0 \\ p\in P}} \tag{1}$$

or its equivalent:

$$\inf_{u_1 \in U} \Im\left(x(t_0), p, u_1\right) \rightarrow \sup_{\substack{|x(t_0)|\le \upsilon \\ p\in P}} \tag{2}$$

We know that the following inequality is fulfilled [9]:

$$\sup_{\substack{|x(t_0)|\le \upsilon \\ p\in P}} \inf_{u_1 \in U} \Im\left(x(t_0), p, u_1\right) \le \inf_{u_1 \in U} \sup_{\substack{|x(t_0)|\le \upsilon \\ p\in P}} \Im\left(x(t_0), p, u_1\right) \tag{3}$$

We can therefore consider the following as an excellent grade or evaluation for the stabilization algorithm:

$$\wp = \sup_{\substack{|x(t_0)|\le \upsilon \\ p\in P}} \inf_{u_1 \in U} \Im\left(x(t_0), p, u_1\right) \tag{4}$$

where

$$\Im = \int_0^{t_1} \left[x^T G x + \Delta u^T N \Delta u\right] dt \tag{5}$$

$G > 0$, $N > 0$ - are with $x=A(p)x+B(p)u$, $x(t_0)=c$, $p \in P$.

These matrices G y N are obtained based on the engineer's experience, which will balance the relative importance of the input and state in the cost function that you are trying to optimize. Then, the algorithm will be able to obtain this excellent evaluation and sometimes it will not, depending on whether a seat point exists in the expression of the particular case being analyzed (3). Figure 1 shows how this maximin test is carried out. The control system behaves like a black box, in which we only know the input information $x(t)$ and the output, which would be the control expression u_1.

The inputs in the testing system are the control u, and the system deviations $x(t)$. This block executes the stabilization quality computer tests. It will then be possible to obtain an evaluation of the stabilization algorithm, information about the initial conditions $x(t_0)$ and the perturbations P.

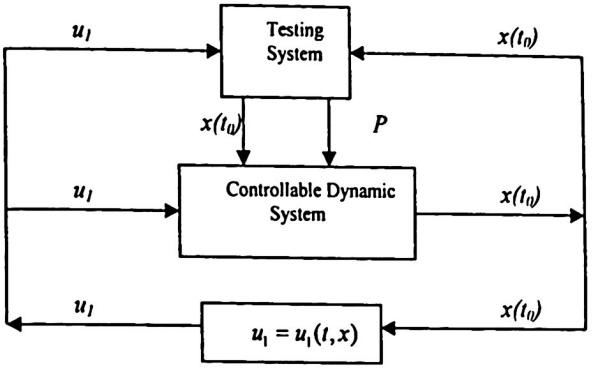


Fig. 1. Diagram of Computer Testing System (CTS).

These tests are organized into three stages:

Stage 1: Solution to the maximin problem: to obtain the excellent solution $\Im^0$; and the solution $x^0(t_0), p^0$ from the maximin problem, which will be used as a strategy in the second stage.

Stage 2: To obtain $\tilde{\Im}$, an estimate of the real evaluation obtained by the stabilization algorithm.

Stage 3: Comparison of $\Im^0$ with $\tilde{\Im}$. The closer the estimate comes to the excellent evaluation, the better the control algorithm will evaluate.

3 Problem Solution

This approach uses an Evolution Strategy (ES) to make Computer Testing System CTS parameters optimization. ES have the following characteristics, desired to parameters optimization of dynamic processes: (a) they are good on real values parameters optimization [5]; (b) usually they are used with self-adaptation of mutation parameters [5]. This is desired because the fitness surface is not known and it can change (in accordance with variations of testing data characteristics); (c) they have low takeover time [5]. This characteristic permit a fast adaptation to the moment characteristics.

This is good for on-line optimization because it is not desired that the population has not became adequate before other changes occur in data characteristics. The used ES utilizes gaussian perturbation mutation with self-adaptation of the standard deviations and one standard deviation for each variable representing an CTS parameter to

be optimized. This kind of mutation is adequate for ordinal numeric parameters optimization.

Therefore, the definitions of the recombination and mutation operations will be given, and the fact that these operations preserve the properties of the positive definite matrices will be demonstrated. The section will conclude with the evolution algorithm.

Definition 1 (Representation and Fitness Evaluation): Search points in ES are symmetric defined positive matrices, that is:

$$\Xi = \left\{ G \mid G \subseteq \Re^n \times \Re^n,\ x^T G x > 0\ \forall x \neq 0\ x \in \Re,\ G = G^T \right\} \tag{6}$$

Given the objective function $\Im : \Xi \to \Re$ (note that $\Im$ is defined as (5), the fitness function Φ is in principle identical to $\Im$, i.e. given an individual $\vec{a} \in I$, we have

$$\Phi(\vec{a}) = \Im(G) \tag{7}$$

Here G is the object variable component of $\vec{a} = (G, \vec{\sigma}, \vec{\alpha}) \in I = \Xi \times A_s$, where

$$\begin{aligned} A_s &= \Re_+^{n_\sigma} \times [-\pi, \pi]^{n_\alpha} \\ n_\sigma &\in \{1, \ldots, n\} \\ n_\alpha &\in \{0, (2n - n_\sigma)(n_\sigma - 1)/2\} \end{aligned} \tag{8}$$

Besides representing the object variable G, each individual may additionally include one up to n different standard deviations σ_i as well as up to $n \cdot (n-1)/2$ rotation angles $\alpha_{ij} \in [-\pi, \pi]$ $(i \in \{1, \ldots, n-1\}, j \in \{i+1, \ldots, n\})$, such that the maximum number of strategy parameters amounts to $\omega = n \cdot (n+1)/2$. For the case $1 < n_\sigma < n$, the standard deviations $\sigma_1, \ldots, \sigma_{n_\sigma - 1}$ are coupled with object variables $G_{1,1}, \ldots, G_{n_\sigma - 1, n_\sigma - 1}$ and σ_{n_σ} is used for the remaining variables $G_{n_\sigma, n_\sigma}, \ldots, G_{n,n}$.

Definition 2 (Mutation): An individual space $I = \Xi \times \Re^n \times \Re^{n \cdot (n-1)/2}$ is assumed. Mutation $m_{\{\tau, \tau', \beta\}} : I^\lambda \to I^\lambda$ is an asexual operator, which yields a triple $\vec{a}' = (G', \vec{\sigma}', \vec{\alpha}')$ when applied to a particular individual $\vec{a} = (G, \vec{\sigma}, \vec{\alpha})$. Then, mutation is formalized as follows

$$\begin{aligned} \sigma_i' &= \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)) \\ \alpha_j' &= \alpha_j + \beta \cdot N_j(0,1) \\ G' &= G + \vec{N}(\vec{0}, C(\vec{\sigma}', \vec{\alpha}')) \end{aligned} \tag{9}$$

The factors τ, τ', and β in equation (9) are rather robust parameters, which Schwefel suggests to set as follows [11]:

$$\begin{aligned} \tau &\propto \left(\sqrt{2\sqrt{n}} \right)^{-1} \\ \tau' &\propto \left(\sqrt{2n} \right)^{-1} \\ \beta &\approx 0.0873 \end{aligned} \tag{10}$$

Usually, the proportionality constants for τ and τ' have the value one, and the value suggested for β (in radians) equals 5°. The $C(\vec{\sigma}',\vec{\alpha}')$ is the covariance matrix with diagonals elements $c_{ii}=\sigma_i^2$, i.e. the variances.

Finally, for mutation of the object variable G the resulting vector $\vec{\sigma}'$ and $\vec{\alpha}'$ are used to create the random object for modifying G, i.e. for its properties of G, it decompose in $G=U\Lambda U^T$, so the mutation operation: $\forall\lambda_i\in\Lambda$ (eigenvalues) with $i=1,2,\ldots,n$, it is carried out $\vec{\lambda}'=\vec{\lambda}+\vec{N}\left(\vec{0},C(\vec{\sigma}',\vec{\alpha}')\right)$ such that $\vec{\lambda}'>0$, this way $G'=U\Lambda'U^T$ where Λ' contains them λ_i for $i=1,2,\ldots,n$, and the new matrix $G'=G+\vec{N}\left(\vec{0},C(\vec{\sigma}',\vec{\alpha}')\right)$ is the result of the mutation that suffer its eigenvalues of the original matrix G.

Definition 3 (Recombination) An individual space $I=\Xi\times\Re^n\times\Re^{n\cdot(n-1)/2}$ is assumed. Recombination $r_{r_s,r_o,r_o}:I^\mu\to I^\lambda$ is an sexual operator. Then, recombination have sexual form, act on two individuals randomly chosen from the parent population, where choosing the same individual twice for creation of one offspring individual is not suppressed. So, we know that G decomposed in $G=U\Lambda U^T$ and $G'=U'\Lambda'U'^T$ is other object variable, then we can express the recombination operator as: $r_{r_s,r_o,r_o}(G_1,G_2)=U_j\Lambda_i U_j^T$ where Λ_i represents an exchange of a subset of n selected elements of a group of $2n$ elements – these elements are $|\Lambda_1\cup\Lambda_2|=2n$, this way $i=1,2,\ldots,\frac{(2n)!}{(2n-n)!}$ and j is alternated by each exchange that is carried out in the group $\{1.2\}$.

Now, when there is obtained a new population of his predecessors, a lot of care must be had if the operations of recombination or mutation did not alter the genetics of the individuals, an option is to verify in each step of the algorithm, that it generates a new population, if there were alterations, another option, which is described in this work, is to demonstrate that these operations do not degenerate the new populations, avoiding the checking in all step of the algorithm and improving the yield of the same one, this is, we improve its computational complexity.

Proposition 1: The operation of recombination preserves the properties of the positive defined matrix.

Proof: Let $X\subset\Xi$ be, for everything $A\in X$, you can express $A=U\Lambda U^T$, and then the rotation $y=U^Tx$ produces the sum of squares

$$\begin{aligned} x^TAx &= x^TU\Lambda U^Tx=y^T\Lambda y \\ &= \lambda_1y_1^2+\lambda_2y_2^2+\ldots+\lambda_ny_n^2 \end{aligned} \tag{11}$$

Now, of the equation (8) those λ_i are substituted for $\lambda_i'>0\ \ \forall i=1,2,\ldots n$

$$\lambda_1' y_1^2 + \lambda_2' y_2^2 + \ldots + \lambda_n' y_n^2 = y^T \Lambda' y = \\ = x^T U \Lambda' U^T x = x^T A' x \tag{12}$$

Where $\lambda_i' > 0$ they are real, then A' it is defined positive. Without loss of generality, one can think that the new values $\lambda_i' > 0$ are the exchanges that are carried out in the operation of crossover to generate to the new matrix C_j for $j = 1,2,\ldots,\frac{(2n)!}{(2n-n)!}$. Reason why, we concludes that the operation of recombination preserve the properties of defined positive symmetrical. ♣

Proposition 2: The operation mutation conserves the properties of the positive defined matrices.

Proof: Let $X \subset \Xi$ be, for everything $A \in X$, then all the values characteristic of the matrix A they are positive, this is $\lambda_i > 0$, for $\forall i = 1,2,\ldots n$, let us suppose that for each one $\lambda_i > 0$ has the corresponding unitary own vector, then one has that

$$Ax_i = \lambda_i x_i \text{ so that} \\ x_i^T A x_i = x_i^T \lambda_i x_i = \lambda_i \tag{13}$$

Since, $x_i^T x_i = 1$ and $x^T A x > 0 \quad \forall x \neq 0$, it will be valid in particular for the own vector x_i and the quantity $x_i^T A x_i = \lambda_i$. Now, of the equation (10) we added to those λ_i, $\beta_i \in N(0,1)$ this is $\gamma_i = \lambda_i + \beta_i$, such that $\gamma_i > 0$, then one has that

$$0 < \lambda_i + \beta_i = x_i^T (\lambda_i + \beta_i) x_i = x_i^T A' x_i \tag{14}$$

This way, the sum of a normalized aleatory number $\beta_i \in N(0,1)$, doesn't affect the property of defined positive of the matrix A original, or the same thing that the new matrix is defined positive. ♣

A (μ, λ)-strategy is required in order to facilitate extinction of maladapted individuals. Selective pressure may not become too strong, i.e., μ is required to be clearly larger than one. Recombination on strategy parameters is necessary (usually, intermediate recombination gives best results).

Algorithm

$t := 0$

Initialize $P(0) := \{\vec{a}_1(0),\ldots,\vec{a}_\mu(0)\} \in \Xi \times \Re^n \times \Re^{n\cdot(n-1)/2}$

Evaluate $P(0) : \{\Phi(\vec{a}_1(0)),\ldots,\Phi(\vec{a}_\mu(0))\}$ where $\Phi(\vec{a}) = \Im(G)$

While $(\iota(P(t)) \neq true)$ **do**

Recombine: $\vec{a}_k'(t) := r'(P(t)) \quad \forall k \in \{1,\ldots,\lambda\}$

Mutate: $\vec{a}_k''(t) := m'_{\{\tau,\tau',\beta\}}(\vec{a}_k'(t)) \quad \forall k \in \{1,\ldots,\lambda\}$

Evaluate: $P''(t) : \{\Phi(\vec{a}_1''(0)),\ldots,\Phi(\vec{a}_\lambda''(0))\}$

Select: $s_{(\mu,\lambda)}(P''(t))$

$t := t + 1$

od

4 Results

We consider the following mathematical model:

$$\ddot{x} = u_1 \qquad u_1 = -k_1 x_1 - k_2 \dot{x}_1 \tag{15}$$

We wish to solve the problem

$$\max_{x(t_0)\le\upsilon} \int_0^{\infty} \left(x_1^2 + \dot{x}_1^2 + \ddot{x}_1^2\right) dt \to \min_{u_1 \in U} \tag{16}$$

whose physical sense is the following: given the worst initial conditions, we want to minimize the system's deviation, as well as the velocity and acceleration of these deviations.

In this case

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad K^T = (k_1, k_2) \quad x^T = (x_1, \dot{x}_1) \tag{17}$$

The expert specialist chooses the following positive, defined matrix

$$G^T = G = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} \quad N = 1 \tag{18}$$

In this example, the seat point exists, with is fulfilled such that

$$\min_{K\in Q} \max_{x(t_0)\le\upsilon} \Im = \max_{x(t_0)\le\upsilon} \min_{K\in Q} \Im \text{ where} \tag{19}$$

$$\Im = \int_0^{\infty} \left(x_1^2 + \dot{x}_1^2 + \ddot{x}_1^2\right) dt$$

After verifying that System (9) is completely controllable, we proceed to the calculation of internal Problem [2], that is:

$$\min_{K\in Q} \Im = \min_{K\in Q} \int_0^{\infty} \left(x_1^2 + \dot{x}_1^2 + \ddot{x}_1^2\right) dt = x^{oT}(t_0) L_0 x^0(t_0) \tag{20}$$

Therefore, the external problem $\max_{x(t_0)\le\upsilon} x^{oT}(t_0) L_0 x^0(t_0) = \mu_{\max}$ is solved. Thus, an optimum value for the functional $1+\sqrt{3}$ was obtained for the worst initial conditions. The graph is shown below.

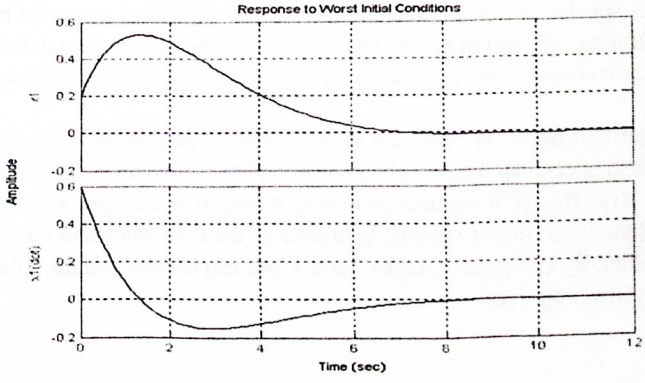


Fig. 2. Asymptotic behavior of the solution when the positive, defined matrix G has been fixed, chosen by a human specialist.

Now we present the result of applying the algorithm based on evolution strategy, in order to automatically choose the best, positive definite matrix G without the intervention of a human expert.

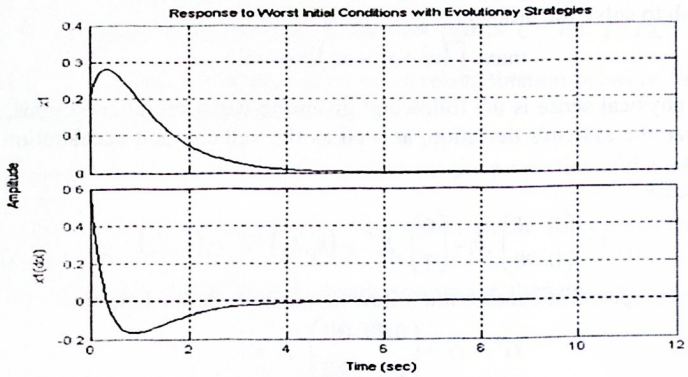


Fig 3: Results of the behavior of the system's coordinates, for the best choice of the pondering matrix, given the worst initial condition.

When the functional does not change in the next generations, this means that we have reached the best individual, that is, the best, positive definite matrix G, which, when applied to the functional, reduces our stabilization time, as is shown in the following graph.

5 Conclusions

The two proposals demonstrate that the recombination and mutation operations preserve the structures of the positive definite matrices. This means that within each iteration of the algorithm, there is no need to verify whether the new generation belongs to its species or not. In other words, the new individual generated by either a mutation or a recombination has the characteristics of being a positive definite matrix. This avoids increasing the complexity in time of the algorithm based on evolution strategies.

The algorithm based on evolution strategy resulted in improving the choice made by the human specialist, thereby obtaining complete autonomy of the computer tests algorithm. Even though it has been proven by means of an example, we can extend its use to include more robust systems by means of the conception and construction of said algorithm. By comparing Graphs 2 and 3, one can observe that the stability of the dynamic system improves.

References

1. N.N. Krasovskii and A.I. Subbotin, Game- Theoretical Control Problems, Springer-Verlag, 1988
2. Alexandrov V. V., Guerra L., Kaleonova I., Trifonova A., Minimax Stabilization and Testing Maximin. University Bulletin, Moscu No.5, 1999
3. Agapie, J. K. Hao, E. Ronald, M. Shoenauer, and D. Snyers, Genetic algorithms: Minimal conditions for convergence. In, Artificial Evolution: Third European Conference; selected papers, pp. 183 - 1943. Berlin Springer 1998
4. Coello, C.A.C.,Christiansen, A.D., and Aguirre, A.H., Using a New GA-Based Multiobjective Optimization Technique for the Design of Robot Arms, Robotica, 16(4), pp. 401-414,1998
5. Eiben, A.E., Smith, J. E. *Introduction to Evolution Computing.* Berlin: Springer, 2003.
6. Coello, C.A.C. and Christiansen, A.D., Multiobjetive optimization of trusses using genetic algorithms, Computers and Structures, 75(6), pp. 647-660, 2000
7. Alexandrov V. V., Guerra L., Sobolevskaya I., Trifonova A., Vargas H., Control Algorithms in Mechatronics, Mathematical Modeling of complex Information Processing Systems, Moscow University Press, 2001
8. Alexandrov V. V., Guerra L., Vargas H., Optimization and Computer Aided Testing of Stabilization Precision, II International Congress of Numerical Methods in Engineering and Applied Sciences, pp. 389-399, January 2002
9. Demyanov V. F. and Malozemov V.M., Introduction to Minimax, Nauka Editorial, Moscu, 1972
10. Eiben, A.E., Smith, J. E. *Introduction to Evolution Computing.* Berlin: Springer, 2003
11. H. P. Schwefel, Numerische Optimierung von Computer-modellen mittels der Evolutionsstrategie, volume 26 of Interdisciplinary Systems Research. Birkhäuser, Basel, 1977

Self-Tuning Neural Controller and its Application to a Non Linear System.

Alfonso Gomez[1], Alfonso Noriega[1], Alberto Aguado[2], and Roberto Salas[3]

[1] Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro, Qro.., México
[2] ICIMAF, La Habana, Cuba
[3] MIDE, Querétaro, Qro., México
{Alfonso Noriega, anoriega@uaq.mx}

Abstract - The performance of an industrial process control system equipped with a conventional controller may be degraded severely by a long system-time delay, dead zone and/or saturation of actuator mechanisms, model and/or parameter uncertainties, and process noises. Recently it has been demonstrated that Artificial Neural Networks can be used to control Non Linear Systems and also it was proposed that a Self-tuning Neural Network implementation is possible with a real time permanent adjustment of the weighting coefficients based on the control error. In the present paper, the Self-tuning Neural Network practical implementation results in a Non Linear Systems are shown and a methodology to adjust the parameters is proposed. As part of the proposal we present a practical method to accelerate net learning by bounding neuron outputs and a time variable learning coefficient.

Keywords: *Neural Networks, Back-propagation, Non Linear Systems, Adaptive Control.*

I. Introduction

The best developed aspect of the mathematical systems theory, deals with the analysis and synthesis of dynamical systems, using well established techniques based on linear algebra, complex variable theory, and the theory of ordinary linear differential equations. Since design techniques for dynamical systems are closely related to their stability properties and since necessary and sufficient conditions for the stability of linear time-invariant systems have been generated, well-known design methods have been established for such systems. In contrast to this, the stability of nonlinear systems can be established for the most part only on a system-by system basis and hence it is not surprising that design procedures that simultaneously meet the requirements of stability, robustness, and good dynamical response are not currently available for large classes of such systems.

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 199-211

Narendra and Parthasarathy (1990), demonstrated that neural networks can be effectively used for nonlinear dynamical systems identification and control. In that paper, direct and indirect control approaches are discussed, explaining that at that moment, methods for directly adjusting the parameters based on the output error (between the plant and the reference model output) were not available. Based on low order nonlinear dynamic systems simulation studies, suggested that identification and control using neural network controllers can find wide application in many areas of applied science.

Later , Cui and Shin (1993) propose a direct adaptive controller and coordinator using neural networks. One of the key problems in designing such a controller/coordinator is to develop an efficient training algorithm. A neural network is usually trained using the output errors of the network, not the controlled plant. However, when a neural network is used to directly control a plant, the output errors of the network are unknown, because the desired control actions are unknown. A simple training algorithm is proposed that enables the neural network to be trained with the output errors of the controlled plant. The only a priori knowledge of the controlled plant is the direction of its output response. However, in that paper it is not discussed the influence of some training parameters, particularly the learning coefficient, over the closed loop dynamics. It is also not remarked that practically with a direct neural control scheme the training stage can be substituted by a permanent and real time adaptation of the weighting coefficients of the neural network.

Recently, Noriega et al (2004), propose a self-tuning neural regulator, inspired in the ideas of Cui and Shin, but with the particular feature that the previous training is substituted by a permanent adjustment of the weighting coefficients based on the control error. After an exhaustive simulation test of the algorithm in several non-linear systems they conclude that a previous training of the network is not generally required; the dynamical performance of the closed loop depends exclusively on the learning coefficient magnitude and they found very convenient to use a variable learning coefficient, using an expression as: $\eta' = \eta + \alpha \text{abs}(e_y)$, so the system present a fast response when the errors are big and then to go slowly to the reference value.

In the present paper, the Self-tuning Neural Network practical implementation results, in a Non Linear Systems are shown and a methodology to adjust the parameters is proposed. As part of the proposal we present a practical method to accelerate net learning by bounding neuron outputs and a time variable learning coefficient. In Section II, the self-tuning neural controller structure is presented. The weighting coefficients adaptation algorithm is developed in Section III and some recommendations to accelerate the network learning (by bounding neuron outputs and using a time variable learning coefficient) are given. Section IV presents the experimental results in a real non-linear system. The paper concludes with Section V, where some directions are given for further development.

II. Self Tuning Neural Controller Structure

In Figure 1, it is shown the scheme proposed by Noriega et al (2004), for the self-tuning neural regulator. The neural network which assumes the regulator function, is a 3 neuron layers perceptron (one hidden layer) which weighting coefficients are adjusted by a modified back-propagation algorithm.

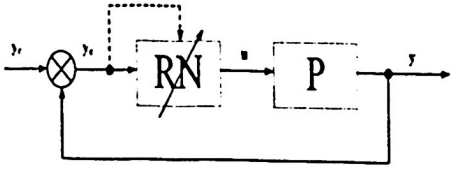


Fig. 1. Self-tuning neural regulator.

In this case, instead of the net output error:

$$e_u(t) = u_d(t) - u(t) \tag{1}$$

it is used the process output error :

$$e_y(t) = y_r(t) - y(t) \tag{2}$$

to adjust the weighting coefficients.

In Fig. 2, it is represented the neural network controller structure. The output layer has only one neuron because by the moment, we are limiting the analysis to one input-one output processes. In the input layer the present and some previous values of the regulation error are introduced, it means that:

$$x(t) = [e_y(t) \quad e_y(t-1) \quad \ldots \quad e_y(t-n)] \tag{3}$$

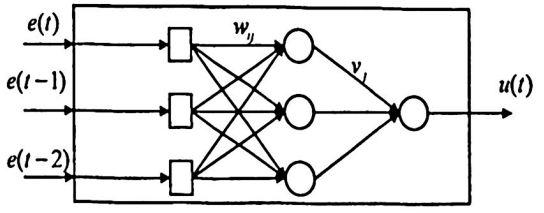


Fig. 2. Neural network controller structure.

For the cases that were simulated in previous works, the value n=2 was sufficient to obtain an adequate closed loop performance. It means that the number of input neurons can be 3. A similar number of hidden layer neurons was equally satisfactory.

III. Weighting Coefficients Adaptation Algorithm

In figure 2, the weighting coefficients w_{ji} and v_j for the hidden layer and output layer input connections respectively, are shown. In what follows, we will detail the adaptation algorithm for those coefficients, which ensures the minimization of a regulation error $e_y(t)$ function.

The output of the j hidden layer neuron may be calculated by means of:

$$h_j = \frac{1}{1+e^{-S_j}}, \quad j = 1,2,3.... \tag{4}$$

Where:

$$S_j = \sum_{i=1}^{3} w_{ji} x_i \tag{5}$$

At the same time, the output layer neuron output value will be:

$$u(t) = \frac{1}{1+e^{-r}} \tag{6}$$

Where:

$$r = \sum_{j=1}^{3} v_j h_j \tag{7}$$

As criteria to be minimized, we defined the function:

$$E(t) = \frac{1}{2} \sum_{k=1}^{t} e_y(k)^2 \tag{8}$$

where it is supposed that the time has been discretized by using an equally-spaced small time interval.

The minimization procedure consists, as it is known, in a movement in the negative gradient direction of the function E(t) with respect to the weighting coefficients v_j and w_{ji}. The E(t) gradient is a multi-dimensional vector whose components are the partial

derivatives $\frac{\partial E(t)}{\partial v_j}$ *and* $\frac{\partial E(t)}{\partial w_{ji}}$, it means that:

$$\nabla E(t) = \begin{bmatrix} \frac{\partial E(t)}{\partial v_j} \\ \frac{\partial E(t)}{\partial w_{ji}} \end{bmatrix} \tag{9}$$

Let us first obtain the partial derivatives with respect to the coefficients of the output neuron. Applying the chain rule, we get:

$$\frac{\partial E(t)}{\partial v_j} = \frac{\partial E(t)}{\partial e_y} * \frac{\partial e_y}{\partial e_u} * \frac{\partial e_u}{\partial u(t)} * \frac{\partial u(t)}{\partial r} * \frac{\partial r}{\partial v_j} \tag{10}$$

$$= e_y . \frac{\partial e_y}{\partial e_u} . (-1) . u(t)(1-u(t)) . h_j \tag{11}$$

In (11) it is used the well known relation:

$$\frac{\partial u(t)}{\partial r} = \frac{\partial\left(\frac{1}{1+e^{-r}}\right)}{\partial r} = \frac{e^{-r}}{\left(1+e^{-r}\right)^2} = \frac{e^{-r}}{1+e^{-r}} \frac{1}{1+e^{-r}} = u(t)(1-u(t)) \tag{12}$$

Let us define:

$$\delta^1 = e_y \, u(t)(1-u(t)) \tag{13}$$

Then:

$$\frac{\partial E(t)}{\partial v_j} = -\delta^1 h_j \frac{\partial e_y}{\partial e_u} \tag{14}$$

In the equation (14), it appears the partial derivative $\frac{\partial e_y}{\partial e_u}$, which can be interpreted as some kind of "equivalent gain" of the process. Further we will make some considerations about that term.

The partial derivative of function E(t) with respect to the weighting coefficients w_{ji}, can be obtained applying again the chain rule:

$$\frac{\partial E(t)}{\partial w_{ji}} = \frac{\partial E(t)}{\partial e_y} * \frac{\partial e_y}{\partial e_u} * \frac{\partial e_u}{\partial u(t)} * \frac{\partial u(t)}{\partial r} * \frac{\partial r}{\partial h_j} * \frac{\partial h_j}{\partial S_j} * \frac{\partial S_j}{\partial w_{ji}} \tag{15}$$

$$= -e_y \frac{\partial e_y}{\partial e_u} u(t)(1-u(t)) v_j h_j (1-h_j) x_i$$

$$\frac{\partial E(t)}{\partial w_{ji}} = -\delta^1 v_j h_j (1-h_j) x_i \frac{\partial e_y}{\partial e_u} \tag{16}$$

Let us define: $$\delta^2{}_j = \delta^1 v_j h_j (1-h_j) \tag{17}$$

And then: $$\frac{\partial E(t)}{\partial w_{ji}} = -\delta^2{}_j x_i \frac{\partial e_y}{\partial e_u} \tag{18}$$

Using equations (14) and (18), the adjustments of weighting coefficients v_j and w_{ji} can be made by means of the expressions :

$$v_j(t+1) = v_j(t) + (\eta \frac{\partial e_y}{\partial e_u}) \delta^1 h_j \tag{19}$$

$$w_{ji}(t+1) = w_{ji}(t) + (\eta \frac{\partial e_y}{\partial e_u}) \delta^2{}_j x_i \tag{20}$$

Where η is the so-called "learning coefficient" and $\frac{\partial e_y}{\partial e_u}$ is the "equivalent gain" of the

The main obstacle to apply the adjustment equations (19) and (20) is that in

general the plant equivalent gain $\frac{\partial e_y}{\partial e_u}$ is unknown. However, in the above mentioned paper by Cui and Shin (1993), it is demonstrated that it is only required to know the sign of that term to ensure the convergence of the weighting coefficients, because the magnitude can be incorporated in the learning coefficient η if the non-restrictive condition $\frac{\partial e_y}{\partial e_u} < \infty$ is accomplished. Besides, the sign of $\frac{\partial e_y}{\partial e_u}$ can be easily estimated by means of a very simple auxiliary experiment, for instance, to apply a step function at the process input.

The assumption that the sign of the gain remains constant in a neighborhood of the operation point of the process is not very strong and it is accomplished in most practical cases. Finally, in the worst of cases, real time estimation of the gain sign could be incorporated, without great difficulties, in the control algorithm. Having in mind the above considerations, the equations (19) and (20) could be written as follows:

$$v_j(t+1) = v_j(t) + \eta \; sign(\frac{\partial e_y}{\partial e_u}) \delta^1 h_j \tag{21}$$

$$w_{ji}(t+1) = w_{ji}(t) + \eta \; sign(\frac{\partial e_y}{\partial e_u}) \delta^2{}_j \, x_i \tag{22}$$

The right value of learning coefficient η can be experimentally determined from the observation of closed loop system performance when some changes are made in the controlled variable set-point. The equations (21) and (22) for the proposed neural controller structure, may be interpreted as the regulator adaptation equations instead of training equations as it is normally done.

III.1 Accelerating the Convergence of the Back-Propagation Algorithm

The back-propagation algorithm as described above encounters the following difficulty. From equations (13), (17), (21) and (22), the increment of weighting coefficients Δv_j and Δw_{ji} can be made by means of the expressions:

$$\Delta v_j = \eta \; sign(\frac{\partial e_y}{\partial e_u}) \delta^1 h_j \qquad \text{Where: } \delta^1 = e_y \, u(t)(1-u(t)) \tag{23}$$

$$\Delta w_{ji} = \eta \, sign(\frac{\partial e_y}{\partial e_u}) \delta^2{}_j \, x_i \qquad \text{Where: } \delta^2{}_j = \delta^1 v_j \, h_j (1 - h_j) \qquad (24)$$

When the actual values h_j and $u(t)$ approach either extreme value, the factors $h_j(1-h_j)$ and $u(t)(1-u(t))$ in equations (23) and (24) make the error signal very small. This implies that an output unit can be maximally wrong without producing a strong error signal with which the weighting coefficients could be significantly adjusted. This retards the search for a minimum in the error. This delay of the convergence is caused by the derivative of the activation function. Unfortunately, any saturating response function is bound to have this property: near the saturation point the derivative vanishes.

On the other hand, when a data acquisition card is used to carry the control signal $u(t)$, the actuator will be saturated for any value of $u(t)$ outside a range determined by the card resolution. For instance, if the card resolution is 12 bits, the actuator will be saturated for $u(t)$ values outside the range [0.00024, 0.99976].

To overcome those difficulties, in the experimental part of this work we normalized the error signal and bounded the weighting coefficients in the range [-4,4], so the algorithm never gets stuck until the local minimum is reached and almost each change in the control signal reflect in the plant behavior.

IV. Experimental Results

The experiment was carried out in a 166MHz Pentium processor computer, a "PCL-818HG" data acquisition card (sample frequency ≤ 100kHz), and a "412" DC servomotor amplifier (24-90 VCD, 10 A). Figure 3 shows a general view of the experiment implementation.

IV.1 System Description

In this system, a pole is connected to the potentiometer shaft, which is used to know the angular position of the pole. A motor with 1600:1 gearbox is connected to the pole with an "L" shaped extension ended in a ring. The ring diameter doubles the pole diameter so there is a backslash in the connection. Depending on the angular position of the motor, for the same motor velocity the pole velocity will change showing a faster response when the ring is connected near the pole base. Figures 4a and 4b show the system at two different positions.

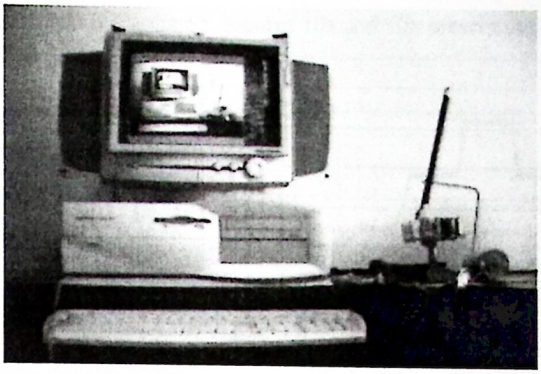


Fig. 3. General view of the experiment implementation.

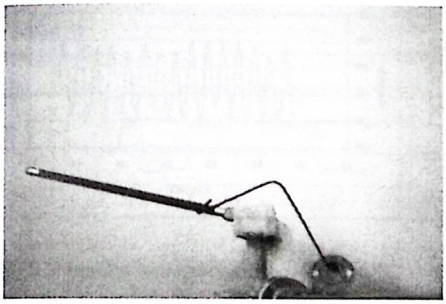


Fig. 4a. System at 160°.

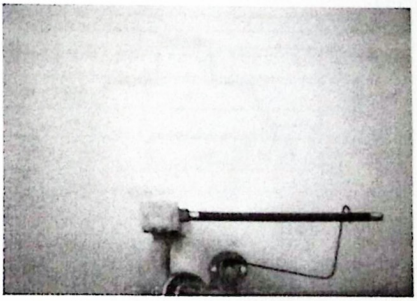


Fig. 4b. System at 0°.

For later reference, a PID controller was tuned to achieve the fastest response of the system without oscillation at step changes of the reference. The controller parameters were adjusted to K = 100, KTτd = 9.2 ms. and (K / Tτi) = (1 / 4.6 s). For those settings, in Figures 5a and 5b the closed loop behavior of the controlled system and the control signal are presented. The required time for the system, to follow the step change of the reference from 23° to 90° is 7 seconds. Any further reduction of this time will lead to oscillation for the step change of the reference from 90° to 157°.

Figure 6a presents the system behavior when the differential control action is eliminated and the figure 6b shows the oscillating response around 157° set point.

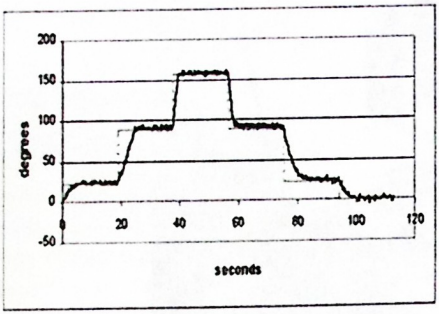


Fig 5a. System controlled whit a PID.

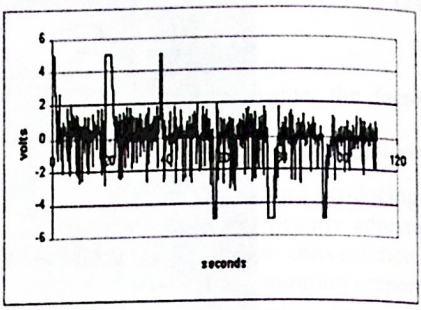


Fig 5b. Control signal for a PID

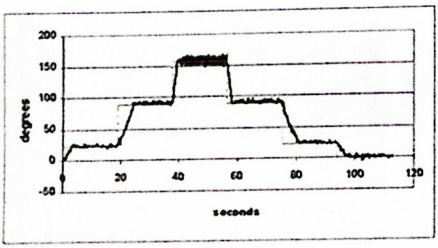


Fig. 6a. System controlled whit a PI.

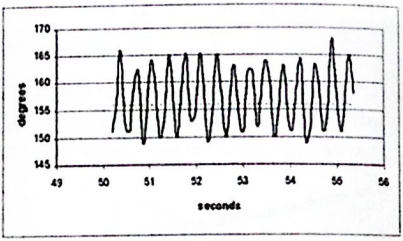


Fig. 6b. Oscillating response around 157°.

The following charts present the experiments whit the self-tuning neural controller. For this experiments the sample and control period were 0.47 ms.

Figure 7 shows the results of the experiments when the learning coefficient is "τ = 0.01". The adaptation time of the weighting coefficients is 50 second and some overshoots show at each reference step. Changing the learning coefficient to "$\tau = 0.01 + 0.1\text{abs}(e_y)$", the adaptation time is reduced to 26 seconds (figure 8).

To reduce the adaptation time it is advisable to use a big learning coefficient at the beginning, and then reduce it to achieve higher stability. Figure 9a shows the results when the learning coefficient $\eta' = \eta + \alpha\text{abs}(e_y)$ is used, and the parameters η and α are adjusted for the first three changes of the reference as follows: (η =0.5 and α=5.0), (η =0.05 and α=0.5) and (η =0.001 and α=0.01). The adaptation time is reduced to 12 seconds and the required time for the system to follow the step change of the reference from 23° to 90° is 3 seconds (43% of the time required for the PID controller in the Figure 5a). The control signal of the self-tuning neural controller, in Figure 9b presents a smoother behavior than

the one shown for the PID in Figure 5b. Figures 10a and 10b present the evolution of two weighting coefficients.

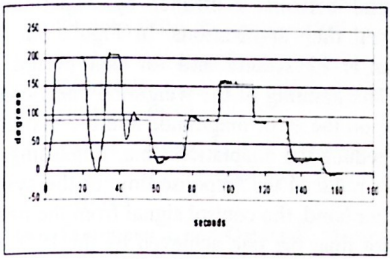


Fig. 7. system controlled whit $\eta = 0.01$.

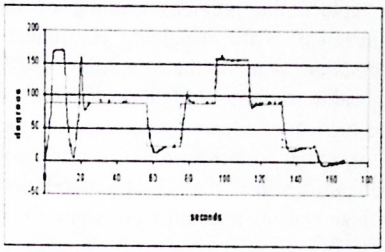


Fig. 8. system controlled whit $\eta = 0.01$ and $\alpha = 0.1$.

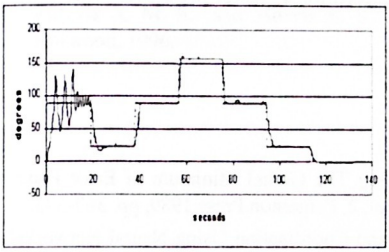


Fig. 9a. η and α adjusted.

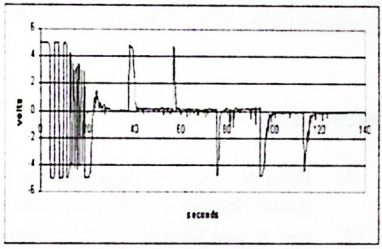


Fig. 9b. Control signal.

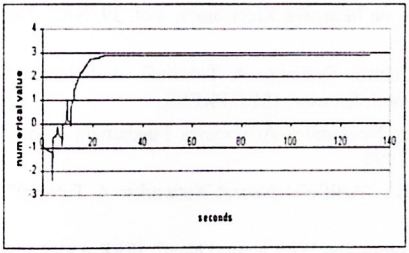


Fig. 10a. Weighting coefficient W[1,2].

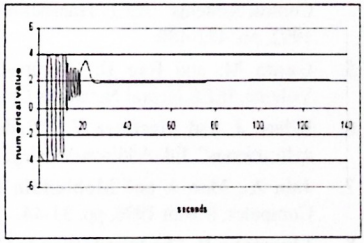


Fig. 10b. Weighting coefficient V[1].

V. Conclusions

According to the experimental results we arise to the following conclusions: It is possible to replace the previous training of the neural network controller by a permanent adjustment of the weighting coefficients in real time applications. In Figure 9a. The adaptation time of the weighting coefficients is 12 second and for that period the algorithm performs 25,000 adaptation cycles. The bonding of the weighting coefficients, and the use of a learning coefficient depending on the error magnitude and the evolution of weighting coefficient played a key role to reduce the adaptation time. Regarding the PID performance, the neural network controller reduced the response time of the system without causing instability problems. On the other hand, the control signal from the neural network controller presents a smoother behavior than the one achieved by the PID, this increases the expected life of the motor and the amplifier by reducing extra heating a vibrations.

To continue with this work it is important to apply the self-tuning neural network controller to other none linear systems including the case of multivariable systems. At the same time it is necessary to implement the algorithm in faster processors like DSP so the adaptation time can be reduced.

References

1. **Baba N.**: "A New Approach for Finding The Global Minimum of Error Function* of Neural Networks". Neural Networks, Vol. 2, Pergamon Press 1989, pp. 367-373.
2. **Cui X.** and **Shin K.**: "Direct Control and Coordination Using Neural Networks". IEEE Transactions on Systems , Man and Cybernetics, Vol. 23, No. 3, May 1993, pp. 686-697.
3. **Freeman J** and **Skapura D.**: "Redes Neuronales Algoritmos, aplicaciones y técnicas de programación". Ed. Addison-Wesley, 1993.
4. **Fukuda T.** and **Shibat T.**: "Theory and Applications for Neural Networks for Industrial Control Systems". IEEE Transactions on Industrial Electronics, Vol. 39, No.6, December 1992, pp. 472-489.
5. **Gupta M.** and **Rao D.**: "Neuro-Control Systems: A Tutorial". A Selected Reprint Volume. IEEE Neural Networks Council, *Sponsor*, IEEE PRESS.
6. **Hilera J.** and **Martínez V.**: "Redes Neuronales Artificiales Fundamentos, modelos y aplicaciones". Ed. Addison-Wesley, 1995.
7. **Jain A., Mao J.** and **Mohiuddin K.**: "Artificial Neural Networks: A Tutorial". IEEE Computer. March 1996, pp. 31-44.
8. **Lippmann R.** "An Introduction to Computing Whit Neural Nets". IEEE ASSP Magazine, April 1987, pp. 4-22.
9. **Moscinski J.** and **Ogonowski Z.**: "Advanced Control with Matlab & Simulink". Ed. Ellis Horwood, 1995.

10. **Narendra K. and Parthasarathy K.**: "Identification and Control of Dynamical Systems Using Neural Networks", IEEE Transactions on Neural Networks, Vol. 1, No. 1, March 1990, pp. 5-27.
11. **Ng G. W.**: "Application of Neural Networks to Adaptive Control of Nonlinear Systems". Ed. John Wiley & Sons Inc, 1997.
12. **Noriega A., Aguado A., Ordaz A.**, and **Rauch V.**: Neural Networks for Self-Tuning Control Systems. Acta Polytechnica Journal of Advanced Engineering. Vol. 44. No 1, 2004, pp. 49-52.
13. **Tai H., Wang J.** and **Ashenayi K.**,: "A Neural Network-Based Tracking Control System". IEEE Transactions on Industrial Electronics, Vol. 39. No.6. December 1992. pp. 504-510.
14. **Tanomaru J.** and **Omatu S.** "Process Control by On-Line Trained Neural Controllers". IEEE Transactions on Industrial Electronics, Vol. 39. No.6. December 1992. pp. 511-521.
15. **Van Ooyen A** and **Nienhuis B.**: "Improving The Convergence of the Back-Propagation Algorithm". Neural Networks, vol. 5, Pergamon Press, 1992. pp. 465-471.
16. **Werbos P.**: "Backpropagation Through Time: What It Does and How to Do It". Proceedings of the IEEE, Vol. 78, no.10, October 1990. pp. 1550-1560.
17. **Zalzala A. M. S.** and **Morris A. S.**: Neural Networks for Robotic Control. Ed. Ellis Horwood, 1996.

Speech Recognition using Based-Phone, Syllables, Triphones and Words on a Digit Corpus in Spanish Language

José Luis Oropeza Rodríguez[1]
Sergio Suárez Guerra[2]

Computing Research Center,
Av. Juan de Dios Batiz esq. Miguel Othon de Mendizabal s/n
Col. Nueva Industrial Vallejo, Mexico, D. F.
[1]joropeza@cic.ipn.mx, [2]ssuarez@cic.ipn.mx,

Abstract. This paper reports results obtained in Automatic Speech Recognition (ASR) using linguistic structural like phonemes, syllables, triphones and words on a digit corpus in Spanish Language. During five decades approximately, the Speech Recognition has been studied by researchers in all the world, simple recognizers have been built, yielding credible performance. But it was soon found that the techniques used in these systems were not easily extensible to more sophisticated systems. For a long time the phoneme has been used because a little amount of them exists. The words also have been utilized but the great quantity of them is a problem when an ASR system is created. The syllable and triphones are the new approaches where this field is focused. In this paper we show the results obtained when we use the continuous and discontinuous speech recognition on a digit corpus, employing the HTK (Hidden Markov Model Toolkit) for that. HTK is a toolkit for building Hidden Markov Models (HMMs). However, HTK was designed for building HMM-based speech recognition tools, in particular recognizers. Thus, much of the infrastructure support in HTK is dedicated to this task. The HTK counts with two major processing stages involved: training tools and recognizer. Firstly, the HTK training tools are used to estimate the parameters of a set of HMMs using training utterances and their associated transcriptions. Secondly, unknown utterances are transcribed using the HTK recognition tools. The results obtained show the capability and difference that exists between the structural linguistic employed for the corpus purposed. We found that the triphones achieve an error rate low in comparison with another structural linguistics proposed in this experiment.

1 Introduction

Automatic Speech Recognition is a recent technology employed in the computer science that allows a computer to identify the words that a person speaks into a microphone or telephone. Despite several decades of research in the area, accuracy greater than 90% is only attained when the task is constrained in some way. For

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 213-226

large-vocabulary speech recognition of different speakers over different channels, accuracy is no greater than 87%, and processing can take hundreds of times real-time.

In English language the use of the triphones has started to try to increment the accuracy and capabilities of the systems mentioned before like Rabiner mentioned in [Rabiner 86]. In the Spanish language, the triphones like a new approach are also interesting. The dominant technology in ASR since 80s is called the HMMs. In recent years, the syllabic structure has been realized incorporating a Knowledge-based system to the training stage. In the training phase an expert system uses ten rules for syllable splitting in Spanish. It receives the energy components STTEF (Short-Term Total Energy Function) and the ERO (Energy Function of the High Frequency) parameters extracted from the speech signal [Oropeza 2005].

In the ASR phoneme-based or another linguistics structure (syllable or triphone), each word in a vocabulary list is specified in terms of its component linguistic. A search procedure is used to determine the sequence of structural linguistic with the highest likelihood. This search is constrained to only look for structural linguistic sequences that correspond to words in the vocabulary list, and the structural linguistic sequence with the highest total likelihood is identified with the word that was spoken.

Thus, the triphone must be studied to know the capability that can offer to the ASR task in the Spanish language and its grammatical structure.

2 Characteristics and Generalities

Speech recognition systems generally assume that the speech signal is a realization of some message encoded as a sequence of one or more symbols. The ASR is constitutive by: training and recognition stages.

The frequency bandwidth of a speech signal is about 16 KHz. However, most of speech energy is under 7 or 7.5 KHz (woman or man voice can change the range mentioned before) dependently. Speech bandwidth is generally reduced in recording. A speech signal is called *ortophonic* if all the spectral components over 16 KHz are discarded. A telephonic lower quality signal is obtained whenever a signal does not have energy out of the band 300-3400 Hz. Therefore, digital speech processing is usually performed by a frequency sampling ranging between 8000 samples/sec and 3200 samples/sec. These values correspond to a bandwidth of 4 KHz and 16 KHz respectively.

Voice is a static procedure that can to have a duration time between 80-200 ms. a simple but effective mathematical model of the physiological voice production process is the excitation and vocal tract model.

The excitation signal is assumed periodic with a period equal to the pitch for vowels and other voiced sounds, while for unvoiced consonants, the excitation is assumed

white noise, i.e. a random signal without dominant frequencies. The excitation signal is subject to spectral modifications while it passes through the vocal tract that has an acoustic effect equivalent to linear time invariant filtering. The model is relevant because, for each type of excitation, a phoneme (or another structural linguistic) is identified mainly by considering the shape of the vocal tract. Therefore, the vocal tract configuration can be estimated by identifying the filtering performed by the tract vocal on the excitation. Introducing the power spectrum of the signal $P_x(\omega)$, of the excitation $P_v(\omega)$ and the spectrum of the vocal tract filter $P_h(\omega)$, we have:

$$P_x(\omega) = P_v(\omega)P_h(\omega) \qquad [1]$$

The speech signal (continuous, discontinuous or isolated) is first converted to a sequence of equally spaced discrete parameter vectors. This sequence of parameter vectors is assumed to form an exact representation of the speech waveform on the basis that for the duration covered by a single vector (typically 10-25 ms) the speech waveform can be regarded as being stationary. Although it is not strictly true, it is a reasonable approximation. Typical parametric representations in common use are smoothed spectra or linear predictive coefficients plus various other representations derived from these. The training stage is involved in the extraction of the parameters mentioned. In our experiments we used the following block diagram for the isolated speech recognition (fig. 1).

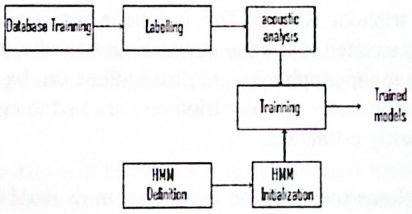


Fig. 1 Block diagram of training step for ASR for isolated words

The database employed consists of the ten digits (0-9) for the Spanish language. Many of the operations performed by HTK which involve speech data assumes that the speech is divided into segments and each segment has a name or label. The set of labels associated with the speech data will be the same as corresponding speech file but a different extension.

Often the final stage of data preparation is to parameterize the raw speech waveforms into sequences of features vectors. In this experiment, Mel Frequency Cepstral Coefficients (MFCCs), which are derived from FFT-based log spectral, will be used.

As figure 2 shows. In HTK is possible to create a set of files to define HMMs structures of each unit employed for the speech recognition. Often, the first step in

HMM training is to define a prototype model. The parameters of this model are not important; its purpose is to define the model topology.

When we employed different linguistic structural to word (phoneme, syllable or triphon), the block diagram changed and it is shown in figure 2.

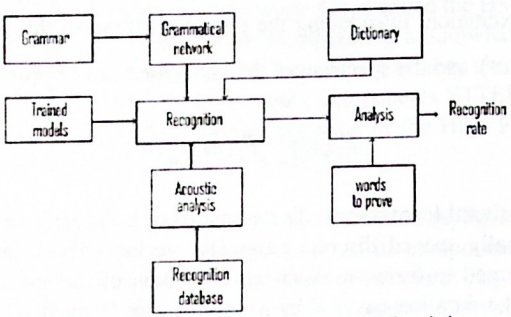


Fig. 2 Training step for ASR using structural linguistic

The new blocks added, in comparison with the block diagram showed in figure 1, represent the integration of the linguistic units mentioned above. Figure 2 shows blocks focused to the phonemes, if the linguistic unit is not that, only it will be changed for the linguistic unit selected.

Given a set of monophone HMMs, the final stage of model building is to create context-dependent triphone HMMs. This is done in two steps. Firstly, the monophone transcriptions are converted to triphone transcriptions and a set of triphone models are created by copying the monophones and distributions can be robustly estimated. Secondly, similar acoustic states of these triphones are tied to ensure that all state distributions can be robustly estimated.

The style of triphone transcription is referred to as word internal. Note that some biphones were generated as contexts at word boundaries were sometimes only include two phones. You can see the alignment data stage and the linking states showed in the figure 2.

In the recognition stage, the block diagram was the following (figure 3).

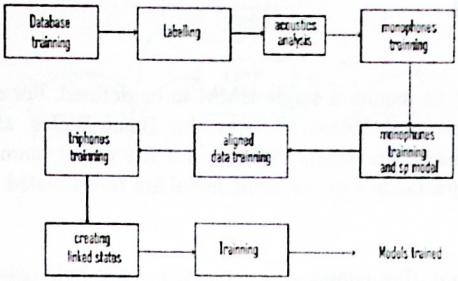


Fig. 3 Block diagram speech recognition used in HTK

As we know, the first stage or any recognizer development project is data preparation. Speech data is needed both for training and for testing. In this case all speech signals will be recorded from scratch and to do scripts are needed to prompt each sentence. In the case of the test data, these prompt scripts will also provided the reference transcriptions against which the recognizer's performance can be measured and a convenient way to create them is to use the task grammar as a random generator. In the case of the training data, the prompt scripts will be used in conjunction with a pronunciation dictionary to provide the initial phone level transcriptions needed to start the HMM training process.

3 Image Acquisition, processing and Segmentation

The database described in this paper counts with a set of forty three utterances for each word (0-9), in total 200 utterances. We used 100 for the training task and the rest for data test.

The training and test data will be recorded using a sound recorder provided with Microsoft Windows. HTK counts with HSLab, this is a program with graphical interface that permits a combined waveform recording and labeling a speech signal. The speech signal has the following characteristics:

- 8 bits per sample
- 11025 frequency sample
- PCM format
- High quality without noise integrated, samples providing from laboratory

HMMs implemented have the following structure:

- 6 states for isolated speech recognition
- 5 states for continuous speech recognition
- Left-right structure
- One Gaussian mixture per state

3.1 Trainning

Some HTK tools require a single HMM to be defined. For example, the isolated-unit re-estimation tool HRest (it uses the Baum-Welch algorithm for the re-estimation). A model previously defined in a file called hmmdef needed to be defined. Then the parameters of the input signal are re-estimated using the speech data files s1, s2, etc.

HMM definition files consist of a sequence of symbols representing the elements of a simple language (HTK uses HInit tool to initialize HMM [Baum 96]). HRest for isolated words and HERest for connected structures are the final tools in the set designed to manipulate isolated unit HMMs. Its operation is very similar to HInit (tool employed to predefine HMM). It expects the input HMM definition to have been initialized and it uses Baum-Welch re-estimation in place of Viterbi training. This involves finding the probability of being in each state at time frame using the Forward-Backward algorithm. This probability is then used to form weighted averages for the HMM parameters.

3.1.1 Dictionary and grammatical structures employed

Before to create a HMM, we must to create a set of word models and define a basic architecture for the recognizer. HTK provides a grammar definition language for specifying simple task grammars. It consists of a set variable definitions followed by a regular expression describing the words to recognize.

The first step in building a dictionary is to create a sorted list of the required words. For this experiment it is quite easy to create a list of required words by hand. For this experiment, the recognizer must handle digit strings only. Examples of typical inputs might be:

- Isolated words: one, zero, three, etc. (one per one not consecutively)
- Connected words: one one two four six seven; two four five zero nine eight, etc.

For the isolated words, a suitable grammar might be:

$digit = ZERO | ONE | TWO | THREE | FOUR | FIVE | SIX | SEVEN | EIGHT | NINE

(SENT-STAR (sil <$digit> sil) SENT-END)

The grammatical network for the experiment is showed in the figure 4.

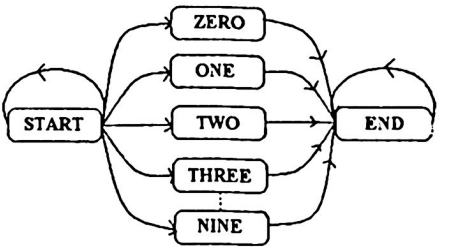


Fig. 4 Grammatical digit corpus structure for isolated speech recognition

When we employed connected digits, the grammatical network looking for such as is illustrated in the figure 5.

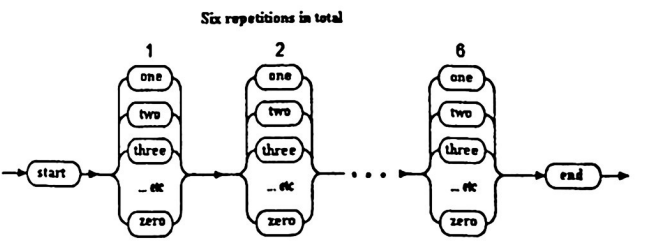


Fig. 5 Grammatical digit corpus structure for connected speech recognition

3.1.2 Trainning Hidden Markov Models

The training and test data can be recorded using the HTK tool HSLab. The training sentences employed in this experiment was obtained from a database of our laboratory, HSLab was used to labeling this training utterances. When HSLab is invoked, a window with a waveform display area in the upper half and a row of buttons appears. When the name of a normal file is given as argument, HSLab displays its contents. To train a set of HMMs, every file of training data must have an associated phone level transcription. In the following figure 6 we can see the labeling realized on a speech signal in the connected words experiment.

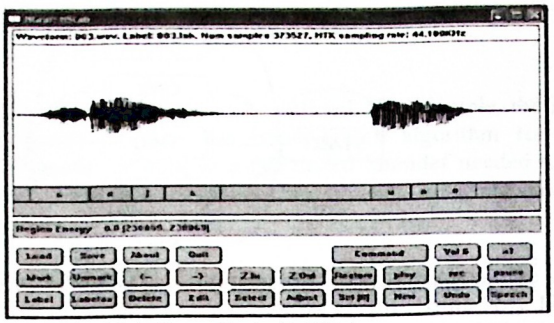


Fig. 6 Labeling realized with HSLab

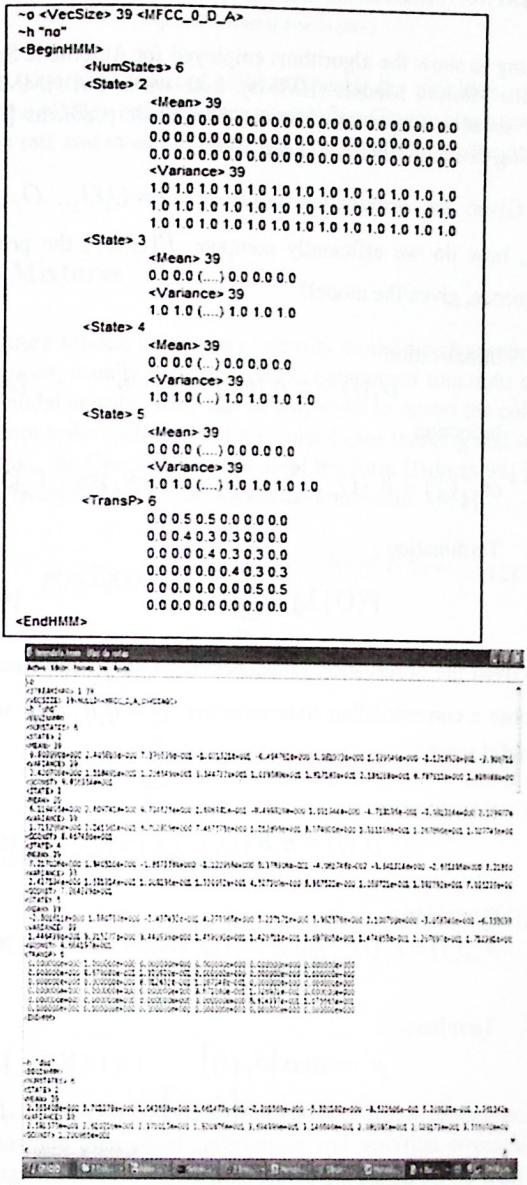
3.1.3 Feature extraction of the speech signal

The final stage of data preparation is to parameterize the raw speech waveforms into sequences of feature vectors. HTK supports both FFT-based and LPC-based analysis. Here Mel Frequency Cepstral Coefficients (MFCCs), which are derived from FFT-based log spectra, were used. Also, we use the tool HCOPY to automatically convert its input into MFCC vectors.

We use W_0 as the energy component, the frame period is 10msec (HTK uses units of 100ns). The FFT used a Hamming Window and the signal had first order preemphasis applied using a coefficient of 0.97. The filter bank was constituted by 26 channels and 12 MFCC coefficients. Then, the system utilized 39 parameters calculated from the length of the parameterized static vector MFCC plus the delta coefficients plus the acceleration coefficients, to represent the speech signal in both: training and test data steps.

3.1.4 Definition of the Hidden Markov Models

The figure 7 shows definition of Hidden Markov Models employed during isolated word recognition. At left the HMM initialized and right the HMM re-estimated.


~o <VecSize> 39 <MFCC_0_D_A>
~h "no"
<BeginHMM>
<NumStates> 6
<State> 2
<Mean> 39
0.0 0.0 0.0

4 Hidden Markov Models

Now, we are going to show the algorithms employed for Automatic Speech Recognition using Hidden Markov Models (HMMs). Like we know, HMMs mathematical tool applied for speech recognition presents three basic problems [Oropeza, 2000] [Rabiner and Biing-Hwang, 1993] y [Zhang 1999]:

Problem 1. Given the observation sequence $O = O_1 O_2 .... O_T$, and a model $\lambda = (A, B, \pi)$, how do we efficiently compute $P(O \lambda)$, the probability of the observation sequence, given the model?

1. Initialization

$$\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \le i \le N \qquad [2]$$

2. Induction

$$\alpha_{t+1}(j) = b_j(O_{t+1}) \sum_{i=1}^{N} \alpha_t(i) a_{ij} \quad 1<j\le N,\ 1\le t\le T-1 \qquad [3]$$

3. Termination

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \qquad [4]$$

Problem 2. Given the observation sequence $O = O_1 O_2 .... O_T$ and the model λ, how do we choose a corresponding state sequence $Q = q_1 q_2 ... q_T$ which is optimal in some meaningful sense?

1. Inicialización

$$\delta_1(i) = \pi_i b_i(O_1) \quad 1 \le i \le N \qquad [5]$$

2. Recursión

$$\delta_{t+1}(j) = b_j(O_{t+1}) \left[\max_{1\le i\le N} \delta_t(i) a_{ij} \right] \quad 1\le j\le N, 1\le t\le T-1 \qquad [6]$$

3. Terminación

$$p^* = \max[\delta_T(i)] \quad 1 \le i \le N \qquad [7]$$

$$q^* = \arg\max[\delta_T(i)] \quad 1 \le i \le N \qquad [8]$$

Problem 3. How do we adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O \lambda)$?

$$a_{ij} = \frac{\text{exp ected number of times from state } s_i \text{ to state } s_j}{\text{exp ected number of transition s from state } s_i} \qquad [9]$$

$$b_{jk} = \frac{\text{expected number number of times in } s_j \text{ and observating } v_k}{\text{expected number of times in state } j} \qquad [10]$$

Then, HMMs algorithms must to solve efficiently the problems mentioned above. For each state, the HMMs can use since one to five Gaussian mixtures both to reach high recognition rate and modeling vocal tract configuration in the Automatic Speech Recognition.

5 Gaussian Mixtures

Gaussian Mixture Models are a type of density model which comprise a number of component functions, usually Gaussian. These component functions are combined to provide a multimodal density. They can be employed to model the colors of an object in order to perform tasks such as real-time color-based tracking and segmentation. In speech recognition, the Gaussian mixture is of the form [Bilmes 98] [Resch, 2001a], [Resch, 2001b], [Kamakshi et al., 2002] and [Mermelstein, 1975].

:

$$g(\mu,\Sigma)(x) = \frac{1}{\sqrt{2\pi^d}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \qquad [11]$$

Equation 12 shows a set of Gaussian mixtures:

$$gm(x) = \sum_{k=1}^{K} w_k * g(\mu_k, \Sigma_k)(x) \qquad [12]$$

In 12, the summarize of the weights give us

$$\sum_{i=1}^{K} w_i = 1 \quad \forall \quad i \in \{1,.....,K\} \; : w_i \geq 0 \qquad [13]$$

6 Results

HTK counts with the tool HResults, which uses dynamic programming to align the two transcriptions and then count substitution and insertion errors. Figure 8 shows that the percentage obtained was 95% for isolated words.

```
hresults -A -T 1 -e ??? sil lista_hmm.txt rec.mlf
====================== HTK Results Analysis =======================
  Date: Wed Jul 05 17:53:48 2006
  Ref : 
  Rec : rec.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=95.00 [H=95, S=5, N=100]
WORD: %Corr=95.00, Acc=95.00 [H=95, D=0, S=5, I=0, N=100]
===================================================================
```

Fig. 8 Digit isolated results using words

And for the connected words the results are shown in figure 9.

```
------------------------ Overall Results ------------------------
SENT: %Correct=60.87 [H=14, S=9, N=23]
WORD: %Corr=92.75, Acc=92.03 [H=128, D=1, S=9, I=1, N=138]
-----------------------------------------------------------------
```

Fig. 9 Digit Connected results using words, format HTK

The line starting with SENT; indicates that of the 23 test sentence utterances (with 6 digits each one, a total of 138), 14 (60.87%) were correctly recognized (discontinuous speech recognition). The following line starting with WORD: gives the world level statistics and indicates that the 138 words in total, 92.75% were recognized correctly. There were 1 deletion error (D), 9 substitution error (S) and 1 insertion error (I).

The use of phone-based (monophones) with the corpus data used, we obtained the following results:

```
hresults -A -T 1 -e ??? sil lista_hmm.txt rec.mlf
------------------------ HTK Results Analysis -----------------------
  Date: Wed Jul 20 11:10:58 2006
  Ref :
  Rec : rec.mlf
------------------------ Overall Results ----------------------------
SENT: %Correct=73.91 [H=17, S=6, N=23]
WORD: %Corr=97.03, Acc=97.03 [H=597, D=8, S=9, I=17, N=573]
---------------------------------------------------------------------
```

Fig. 10 Connected digits application using phonemes as linguistic structure

Context-dependent triphones can be made by simple cloning monophones and then re-estimating using triphone transcriptions. HTK counts with HLEd tool to do that. Figure 11 shows the results:

```
hresults -A -T 1 -e ??? sil lista_hmm.txt rec.mlf
------------------------ HTK Results Analysis -----------------------
  Date: Wed Jul 17 12:13:48 2006
  Ref :
  Rec : rec.mlf
------------------------ Overall Results ----------------------------
SENT: %Correct=60.87 [H=14, S=9, N=23]
WORD: %Corr=96.76, Acc=93.52 [H=597, D=10, S=10, I=20, N=617]
---------------------------------------------------------------------
```

Fig. 11 Connected digits application using triphones linguistic structure

From [Oropeza 2005] a 98% was reported using syllables. The results show the comparison between different linguistic structural in the corpus digit for Spanish language.

7 Conclusions and future works

The main feature that is used to characterize the complexity of speech recognition is weather the speech is connected or is spoken one word at time. In connected speech, it is difficult to determine where one word ends and another begins, and the

characteristics acoustic patterns of words exhibit much greater variability depending on the context. Isolated word recognition systems do not have these problems since words are separated by pauses.

HTK is a set of tools that permit to create an Automatic Speech Recognition (ASR) secure and efficient. It permits to use words, phoneme or triphone based on monophoneme definition as linguistic structures. Also, it contains with a graphical tool that permits to create labels for each word. The rest of tools are accessed via DOS shell prompt.

In this paper we show the results obtained in both Isolated and connected speech recognition systems using HTK for a digit corpus in Spanish language. At the same time, we employed three structural linguistic (words, phonemes and triphone) and we compared them with the results obtained when the syllables were used for the same corpus without employ HTK.

The differences found it between the accuracy recognition using different linguist structures is a consequence of so much things. Firstly, the labeling proposed here was effected using HSLab tool, which does not permit an automatic splitting of the word. Then, human errors are possible to appear and it affects a good response of the system. It explains because the error rate is lower than compared with the experiment using syllable structures, where a Knowledge based system was used for the splitting the word.

So, the result obtained shows an important accuracy to consider an adequate real implementation of them. The most important result extracted from the experiment is the integration of triphone structure and its repercussions for future works. The possibilities to use other linguistic structures in ASR increment the faculty to create real systems using new approaches. Also, incorporating the analysis previously realized in automatic speech recognition with noise can be another way to consolidate the study of linguistic structures.

References

1. Pope A. R. "Model-Based Object Recognition. A survey of recent research", University of British Columbia, Vancouver, Canada, Technical Report 94-04, January 1994.

2. Singh S., Markou M., Haddon J., "Detection of new image objects in video sequences using neural networks", Proc. SPIE Vol. 3962, p. 204-213, Applications of Artificial Neural Networks in Image Processing V, Nasser M. Nasrabadi; Aggelos K. Katssagelos; Eds., 2000.

3. Fay R., Kaufmann U., Schwenker F., Palm G., "Learning objects recognition in a neurobotic system". In: Horst-Michael Grob, Klaus Debes, Hans-Joachim Böhme (Eds.) 3rd Workshop on Self Organization of AdaptiVE Behavior (SOAVE 2004). Fortschritt-Berichte VDI, Reihe 10 Informatik / Kummunikation, Nr. 743, pp. 198-209, VDI Verlag, Düsseldorf, 2004.

4. Wang W., Zhang A. and Song Y., "Identification of objects from image regions", IEEE International Conference on Multimedia and Exp (ICME 2003), Baltimore, July 6-9, 2003.
5. Vision Components, Technical Documentation VCM40, VCM50, Vision Components GmbH Ettlingen, Germany, Oktober 2003.
6. Abhijit S. Pandya, Robert B. Macy, Pattern Recognition with Neural Networks in C++, edit. CRC PRESS & IEEE PRESS. A CRC Cook Published in Cooperation with IEEE PRESS, 1996.
7. Felzenswalb P. and Huttenlocher D., "Efficiently computing a good segmentation". In IEEE Conference on Multimedia and Expo (ICME 2003), Baltimore, July 6-9, 2003.
8. Bishop C. M., Neural networks for Pattern Recognition, Oxford University Press, 1995.
9. Nicolas Amezquita Gomez and Rene Alquezar., "Object Recognition in Indoor Sequences by Classifying Image Segmetnation Regions Using Neural Networks, 10th Iberoamerican Congreso on Pattern Reognition, CIARP 2005, Havana, Cuba, November 2005, Proceedings.

Real-time Fuzzy Digital Filter: Basic Concepts.

[1] J. C. García Infante, [1,2] J. J. Medel Juárez, [1] P. Guevara López.
[1] Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada
(CICATA-IPN UNIDAD LEGARIA).
[2] Centro de Investigación en Computación
(CIC-IPN UNIDAD ZACATENCO)

Abstract. This work presents an analysis of fuzzy logic (to see: [1], [3], [8], [9], [10], [11]) applied to digital filters (to see: [2], [4], [8] and [12]) and real-time systems (to see: [15] and references in it), for to do a concept of real time fuzzy digital filters. We will describe the basic characteristics of operation and meaning of this kind of filters in order to know how the signals will be classified by the filter (to see: [12] and [15]), we resolve how to construct and characterize the membership function [3] and the knowledge base in a probabilistic manner [12], making a description of the Real-time conditions that the Real-time Fuzzy Digital Filter (RTFDF) has perform: the basic results are described in formal sense using the basic definitions considering the concepts develop in others papers referenced in this; finally this work shows a simulation of the RTFDF operation.

Keywords: Fuzzy logic, digital filter, adaptive filter, knowledge base, Real-time.

© L. Sánchez, O. Espinosa (Eds.)
Control, Virtual Instrumentation and Digital Systems.
Research in Computing Science 24, 2006, pp. 227-237

1 Introduction

The fuzzy filter system [1], works in loop form and its parameters are adapted in a dynamical form ([2], [4], [6], and [11]) this is based in the error *e(k)* ([5] and [6]) which is the difference between the desired response *d(k)* and the actual signal *y(k)* ([5], [6], [11]).

The criterion is based on the pursuit error *e(k)* permitting this to generate the membership function (the function which approximates the signal *y(k)* to *d(k)*) for to adjust the parameters of the filter which the value of the error *e(k)* should be near to γ [15].

In the fuzzy filtering each rule defined as membership function is limited by a specific sequence given by the response of the system within of the correct operation of the system range and also limited by a distribution function predefined and where the responses of the system are bounded. This concept is described in the follows description:

The fuzzy filter is composed by the next elements considering the concepts studied in [1], [2], [5], [6], [7], [8], [9], [10], and [15]:

1. Rule base, (*IF-THEN*), it contains the quantification of the process by fuzzy logic, realizing an expert linguistic description about how obtain an optimal process filtering, using predefined intervals.
2. Inference Mechanism, this makes the expert decision based on an interpretation and this is applied to know how of what the best way to filter a process by predefined intervals is.
3. Input Fuzzy Inference, this converts the input of the filter into usefulness information to the inference mechanism for to activate and apply the rules.
4. Output Fuzzy Inference, this converts the results or conclusions that the inference mechanism obtains into actualized inputs for the process based in the knowledge base of the system or process that is analyzing.

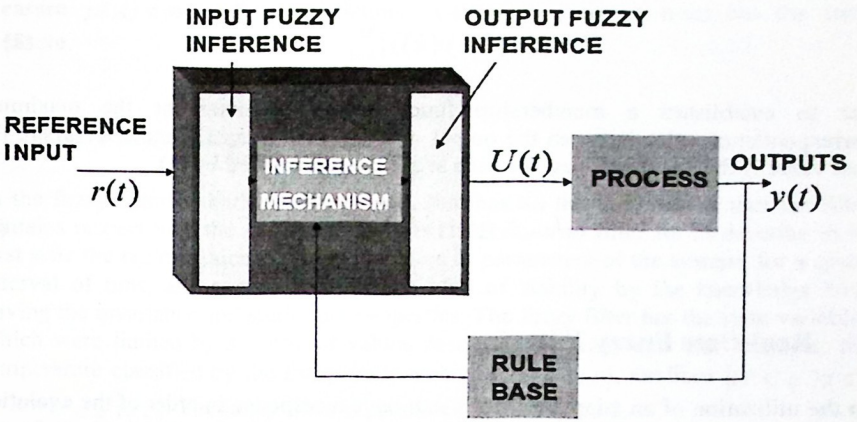


Figure 1. Fuzzy filter process: Description.

2 Membership Functions

DEFINITION: A fuzzy filter is a filter which its error signal ([5] and [15]) is used for to make a membership function ([1], [3], [8], [9], and [10]) required for the adapting algorithm ([2], [4], and [8]).

Here are realized the necessary adjustments for example: regulation of the exciting signals and the filter coefficients using the knowledge base, by this manner we can obtain the minimum previous selection of the filter response with respect to a reference criterion, in a distribution sense.

The most used criterion in the literature as the objective function is the quadratic mean error (QME), described by the second probability moment (to see: [4], [5], and [15]) in a recursive form (1):

$$E_n = \left[\frac{1}{n}\left((n-1)E_{n-1}^2 + e(n)^2\right)\right]^{\frac{1}{2}} \tag{1}$$

In the fuzzy filtering each rule that defines a membership function is limited by a specific sequence given by the response of the system within of the correct operation of the system range and also limited by a distribution function predefined where the responses of the system are limited: the filter will give a specific response *y(k)* respect to the desired signal *d(k)* for each objective function: where each function is limited by an error criterion that establishes a determined action or response of the system (low, medium, high). A set of data is labeled with N desired signals *d(k)* and the respective responses *y(k)* that you can see in (2):

$$T = \{(d(k), y(k))\}_{k=1}^{N} \tag{2}$$

For to establishes a membership function has to determine the maximum correspondence value between the output *y(k)* and the desired signal *d(k)*, where the best value is the minimum cost for each sequence represented by (3).

$$J_{\min} = \min_{i \to N} \{J(d_i, y_i)\} \tag{3}$$

3 Real-time Fuzzy Filtering

In the utilization of an adaptive filter it can have a response in order of the evolution of an specific process for a kind of excitation, but it requires that the loop for the inputs modifies in a optimum form the system responses, without generate a change in its states, and has to operate in Real-time.

The fuzzy filter classifies its different levels of operation by the membership functions, for to give an specific response by the value of the error function J_{min} to a given input; corresponding a specific response for each kind of input limited by the error function that is also limited by γ. The classification of the system response is realized by the fuzzy filter, for to know the type of the inputs or parameters; permitting to know in an indirect way if the system has invariance, the input is stationary and other properties. It establishes the region where the gains or parameters of the system are, describing the estimation state. The application of the fuzzy filter avoids the initial instability of the error function that is limited by an interval within the system operation.

The classification of the membership functions is realized by limiting the criterion (error function), by metric intervals established in each function. Each membership function is established inside the distribution function of the error criterion and could exits diverse membership functions within it, there are characterized within the distribution function inside the system characteristics. This classification is based in the Lebesgue integral because it can give simple conditions permitting that the multiple integrals be expressed in an iterate form and to describe the system in an ordered form, in almost every point.

For to apply the Lebesgue integral (to see: [5], [13], [14] and [15]) it requires to define in each subset of R a function, and it would be associated with a size or measure $\mu(A) \geq 0$, this function must satisfy some properties: For a segment $A = [a,b]$ the measure is given by its length, $\mu(A) = /b - a/$; If A is the union of the sets $\{A_i : i = \overline{1,n}\}$, disjoints by pairs, and $\mu(A) = \sum_{k=1}^{\infty} \mu(A_k)$, If A is a whole with

measure $\mu(A) < \infty$, and its translation $x + A = \{x + y : y \in A\}$ must has the same measure.

4 The Fuzzy Filter Elements

In the fuzzy filter is the knowledge base, that has all the information that the filter contains respect with the system to monitor, pursuit, or to filter for to describe in its first state the region which contains the gains or parameters of the system, for a given interval of time, and can establishes the kind of stability by the knowledge base having the invariance and stationary properties. The fuzzy filter has the state variables which were limited by a range of values described by fuzzy sets (for example, the temperature classified by the fuzzy logic as high $(32^{o}C\ a\ 43^{o}C)$, medium $(17^{o}C\ a\ 32^{o}C)$ or low $(9^{o}C\ a\ 17^{o}C)$. The fuzzy filter is presented in the Figure 2.

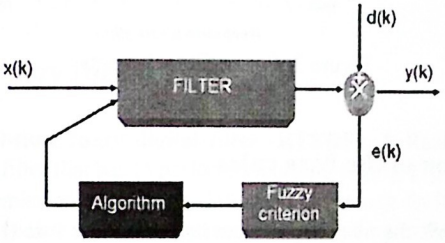


Figure 2. Fuzzy filter

Where $x(k)$ is the linguistic value, $T(k)$ is the set of the representation values that could take $x(k)$ within a limited interval, $d(k)$ is the fuzzy variable of reference, $e(k)$ is the fuzzy value where the set $\{\gamma_i : \gamma_i > 0,\ \forall i \in N\}$ has $inf\{\gamma_i\} \to \lambda,\ \lambda > 0$ and the $sup\{\gamma_i\} \to |\lambda|,\ |\lambda| < 1$, where $\gamma *$ is small in an infinite form because $y(k)$ should be almost equal to $d(k)$, $e(k)$ is the difference between $y(k)$ and $d(k)$, and the criterion contains the values o membership functions for each case. The elements of the fuzzy filter are shown in Figure 3.

1 Introduction

The fuzzy filter system [1], works in loop form and its parameters are adapted in a dynamical form ([2], [4], [6], and [11]) this is based in the error *e(k)* ([5] and [6]) which is the difference between the desired response *d(k)* and the actual signal *y(k)* ([5], [6], [11]).

The criterion is based on the pursuit error *e(k)* permitting this to generate the membership function (the function which approximates the signal *y(k)* to *d(k)*) for to adjust the parameters of the filter which the value of the error *e(k)* should be near to γ [15].

In the fuzzy filtering each rule defined as membership function is limited by a specific sequence given by the response of the system within of the correct operation of the system range and also limited by a distribution function predefined and where the responses of the system are bounded. This concept is described in the follows description:

The fuzzy filter is composed by the next elements considering the concepts studied in [1], [2], [5], [6], [7], [8], [9], [10], and [15]:

1. Rule base, (*IF-THEN*), it contains the quantification of the process by fuzzy logic, realizing an expert linguistic description about how obtain an optimal process filtering, using predefined intervals.
2. Inference Mechanism, this makes the expert decision based on an interpretation and this is applied to know how of what the best way to filter a process by predefined intervals is.
3. Input Fuzzy Inference, this converts the input of the filter into usefulness information to the inference mechanism for to activate and apply the rules.
4. Output Fuzzy Inference, this converts the results or conclusions that the inference mechanism obtains into actualized inputs for the process based in the knowledge base of the system or process that is analyzing.

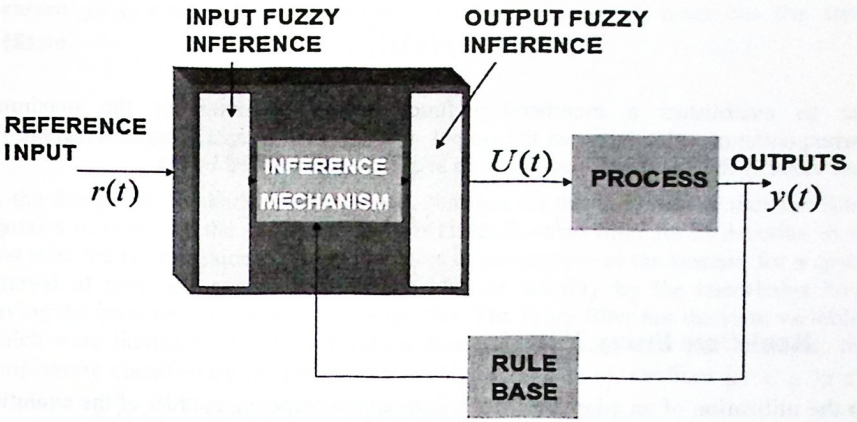


Figure 1. Fuzzy filter process: Description.

2 Membership Functions

DEFINITION: A fuzzy filter is a filter which its error signal ([5] and [15]) is used for to make a membership function ([1], [3], [8], [9], and [10]) required for the adapting algorithm ([2], [4], and [8]).

Here are realized the necessary adjustments for example: regulation of the exciting signals and the filter coefficients using the knowledge base, by this manner we can obtain the minimum previous selection of the filter response with respect to a reference criterion, in a distribution sense.

The most used criterion in the literature as the objective function is the quadratic mean error (QME), described by the second probability moment (to see: [4], [5], and [15]) in a recursive form (1):

$$E_n = \left[\frac{1}{n}\left((n-1)E_{n-1}^2 + e(n)^2\right)\right]^{\frac{1}{2}} \tag{1}$$

In the fuzzy filtering each rule that defines a membership function is limited by a specific sequence given by the response of the system within of the correct operation of the system range and also limited by a distribution function predefined where the responses of the system are limited: the filter will give a specific response *y(k)* respect to the desired signal *d(k)* for each objective function: where each function is limited by an error criterion that establishes a determined action or response of the system (low, medium, high). A set of data is labeled with N desired signals *d(k)* and the respective responses *y(k)* that you can see in (2):

$$\int m^* \, dm = \int (\arg\min \inf_{k \geq m} P\{|\hat{a}_k - a| \leq \Delta\} = 1) dP \subseteq [0 + \beta, 1 - \alpha) \tag{5}$$

Each RTFDF as parameters estimator has a limited error function: where Δ is the limit of the error defined by the variance of the system perturbations, in the fuzzy case this parametric variations within the error limited by the variance represents the membership functions or corresponding criterions for the reference model to each fuzzy variable [11].

Local. Implies the stability of the process trough its parameters $a(k)_l$ actualized at each iteration or at each fuzzy variable and the finalized times $f(k)_l$ of the RTFDF within of the absolute corresponding limits $[Id(k)_{l\ min}, LD(k)_{l\ max})$.

The RTFDF will give a stable local response if the estimation to the set of the SLIT parameters of the filter, being all of them inside of the unit circle for all the k interval considering either convolution with respect to the follows filter. This is:

$$\{a_e(k)\}_l \subseteq [0 + \beta, 1 - \alpha), \quad k = \overline{1,n}. \tag{6}$$

7 Simulation:

For the simulation of the fuzzy filter and to estimate the estates and parameters of the system we were used the Kalman filter [5]. The system was characterized with its operational levels by using fuzzy filtering concepts [11].

The system is described by its estates as:

$$x(k) = a(k)x(k-1) + w(k), \tag{7}$$

And the output is described by:

$$y(k) = x(k) + v(k), \; x(k), \; w(k), \; v(k) \in R^{n \times 1} \tag{8}$$

Where: $x(k)$= is the internal estate of the system, $\{a(k)\}$=are the adjusting parameters, $\{w(k)\}$=are the noises of the system, $y(k)$= is the output of the system, $v(k)$= are the output noises. For to obtain the different levels of operations we must have the mean of the system output $y(k)$, and to get its variance in order to describe the levels of the response (low, medium and high). Figure 4, shows the response levels of the output $y(k)$:

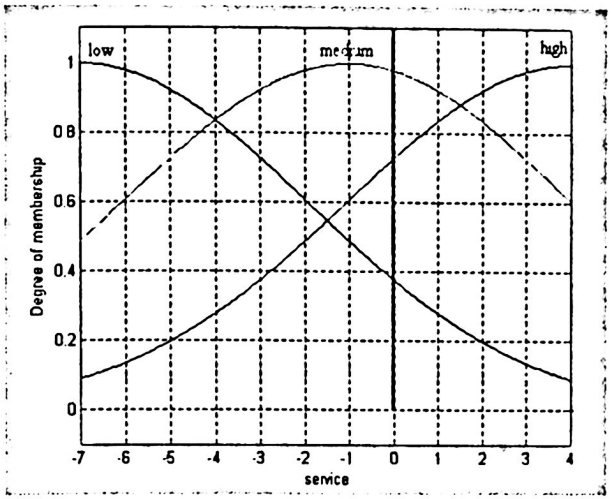


Figure 4. Response levels of *y(k)*

Figure 5, shows the Functional described in (3) with respect to the filter:

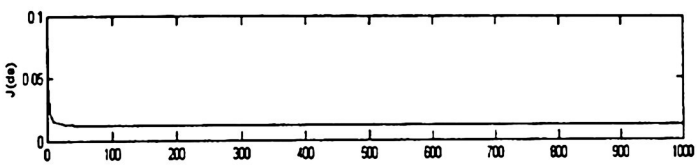


Figure 5. Error functional

Figure 6, shows the Functional Levels in distribution sense, bounded these by the first two probability moments:

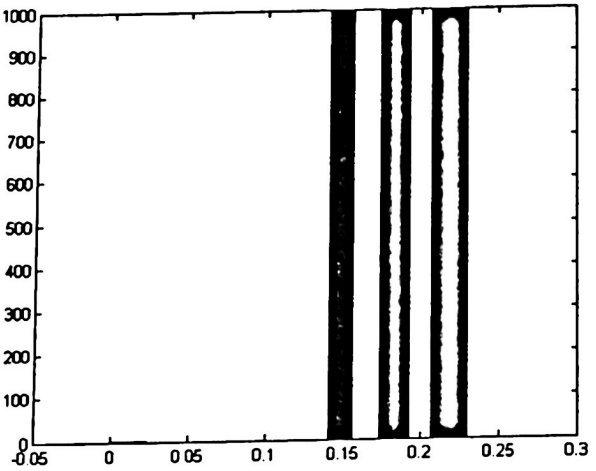


Figure 6. Distribution of error function *J(k)*

Figure 7, shows the Membership Functions classified in probability distribution sense with respect to *J(k)* bounded it by the first two probability moments:

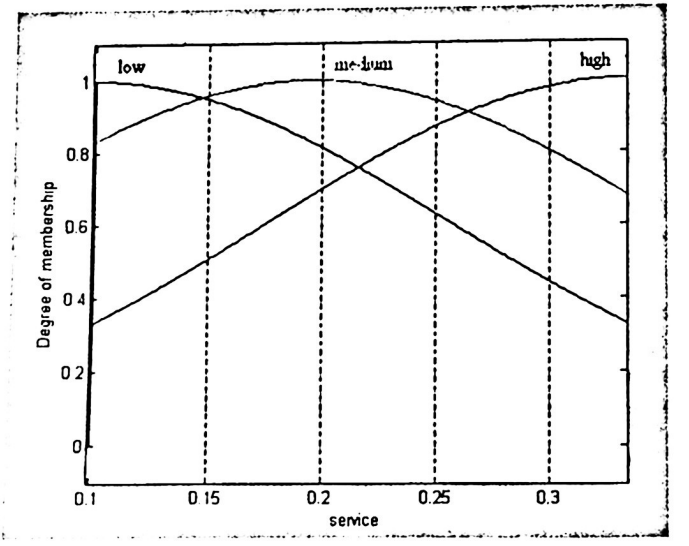


Figure 7. Membership functions *J(k)*

8 Conclusions

This work was presented considering basic concepts of fuzzy logic applied into digital filters with Real-time systems, which made origin to the basic concept of real time fuzzy digital filters, described as RTFDF. We were presented here the basic operation characteristics and meaning of this kind of filters in order to know how the signals was classified by the filter, we resolved how to construct and characterize the Membership Function and the Knowledge Base in a probabilistic manner, making a description of the real time conditions that the RTFDF was performed, and finally this work showed an illustrative simulation of the RTFDF operation.

References

[1] Abdul, S. "Fuzzy logic and its uses", http://www.doc.ic.ac.uk, (2006).

[2] Ali, H. S. "Fundamentals of adaptive filters", Complex systems

[3] Amble, T. "Logic Programming and knowledge engineering", Addison Wesley (1987).

[4] Gustafsson F, "Adaptive Filtering and change detection", John Wiley & Sons, Ltd, (2000).

[5] Haykin, S. "Adaptive Filtering", Prentice Hall (2001).

[6] Jantzen, J. "A tutorial on adaptive fuzzy control", Technical University of Denmark (2002).

[7] Leydesdorff, L. "The construction and globalization of the knowledge base in interhuman communication systems", Canadian Journal of communication, vol 28, (2003) 267-289.

[8] Margaliot, M., and Langholz, G. "New approaches to fuzzy modeling and control design and analysis", World Scientific (2000).

[9] Montejo, M. "Lógica difusa y control difuso", http://www.redeya.com, (2006).

[10] Morales, G. "Introducción a la lógica difusa", Cinvestav-IPN, (2002).

[11] Passino, K. M. "Fuzzi control", Addison Wesley (1998).

[12] Schneider, M., and Kandel, A. "Fuzzy expert systems tools", Wiley (1996)

[13] Sanchez, C. "Integral de Lebesgue-Stieltjes", Marchena, Sevilla (1998).

[14] Manuel, L. "Teoría de la medida e integral de Lebesgue", Universidad Nacional del Rosario (2003).

[15] Medel, J., and Guevara, P. "Caracterización de filtros digitales en tiempo-real para computadoras digitales", Computación y Sistemas, vol. VII, No. 3, (2004).

AUTHOR INDEX

	Pág.
A	
Aguado, Alberto	192
Aldape Pérez, Mario	11
Alexandrov, Vladimir	183
Ali Ghaffari	156
Álvarez, Guillermo	136
C	
Camacho Nieto, Oscar	82
Carbajal Hernández, José J.	91
Cruz Contreras, Agustín	1
D	
Díaz de León Santiago, J. Luis	1
E	
Espinosa Sosa, Osvaldo	82
F	
Fernández, Luis Miguel	136
G	
García Infante, Juan Carlos	219
Gómez, Alfonso	192
González Martínez, Yovany	166
Gorostiaga, Lázaro	166
Guevara López, Pedro	42,101,219
H	
Hernández Zavala, Antonio	82
J	
Jaime Fonseca, Beatriz A,	53
K	
Kaimal M. R.	126
L	
Leyva del Follo, Luis Eduardo	31
Lómas, Víctor	112
López, Mónica	183
López, Luis Octavio	11
López Cárdenas, Rodrigo	53
López Chau, Asdrubal	101
López Silva, Blanca E.	42
M	
Machirán, Miguel	136
Manrique Ramírez, Pablo	1
Mansour Nikkah Bahrami	156
Medel Juárez, José J.	42,101,219
Mejía Álvarez, Pedro	31
N	
Nava, Israel	112
Noriega, Alfonso	192
O	
Oropeza, José Luis	72,205
P	
Perán, José Ramón	166
Pérez Echazabal, Lucrecia	146
Progrebnyak, Oleksiy	1,91
R	
Rajesh R.	126
Rosas, Lorna	183
S	
Salas, Roberto	192
Salcedo, Ante	63
Sánchez, Ileana	136
Sánchez, Luis Alejandro	136,146
Sánchez F., Luis Pastor	91,146
Seyed Mahdiar Ghaffari	156
Silva López, Héctor	21
Suárez Guerra, Sergio	21,72,205
Suresh K. K.	126
Sureshkumar T.	126
T	
Tamayo Fong, Alain	31

	Pag.
V	
Vargas, Héctor	183
Vicente Vivas, Esaú	63,112
W	
Wenbo Xu	173
Wenke Zhao	173
Y	
Yañez Márquez, Cornelio	11
Z	
Zanella, Vittorio	183
Zilei Chai	173

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
Noviembre de 2006.
Printing 500 / Edición 500 ejemplares.

We are privileged to present this special issue that contains the best works presented at the 7th edition of the International Conference on Control, Virtual Instrumentation and Digital Systems CICINDI 2006, which is organized by the Center of Computing Research belonging to National Polytechnic Institute. This year the conference was held in México City. CICINDI is an international conference that joins contributions obtained by specialists in areas such as: Computer Architecture, Digital Signal Processing, Virtual Instrumentation, Digital Systems Design and Control, specialized areas that are having a high impact on contemporary societies. We received several papers from researches of many parts of the world. We accepted only the best papers. Every paper was evaluated by three reviewers with the common objective of select the papers with the highest quality.

We want to express our acknowledgement to all the people that contribute with this magazine. We are specially indebted with reviewers and scientific committee members for their outstanding work and advice. Finally We would also like to thank to the National Polytechnic Institute for facilities given in order to achieve the success of this Conference, which we believe, is so much important for the academic and scientific life in this country.

Luis Pastor Sánchez Fernández
Osvaldo Espinosa Sosa

ISSN: 1870-4069

www.cic.ipn.mx/rcs
www.ipn.mx

70 Aniversario
INSTITUTO POLITÉCNICO NACIONAL
1936 • 2006

INSTITUTO POLITÉCNICO NACIONAL
"La Técnica al Servicio de la Patria"